REPORT RESUMES

ED 003 485 24

EFFECTIVENESS OF THREE DIFFERENT BASAL READING SYSTEMS ON FIRST GRADE READING ACHIEVEMENT.

BY- TANYZER, HAROLD J. ALPERT, HARVEY
HOFSTRA UNIV., HEMPSTEAD, N.Y.
REPORT NUMBER CRP-2720 PUB DATE 65
EDRS PRICE MF-$0.27 HC-$5.52 138P.

DESCRIPTORS- *BASIC READING, GRADE 1, *READING ACHIEVEMENT, ACHIEVEMENT GAINS, *BEGINNING READING, SEX DIFFERENCES, INTELLIGENCE DIFFERENCES, VOCABULARY DEVELOPMENT, SPELLING, ORAL EXPRESSION, WORD RECOGNITION, *READING READINESS, HEMPSTEAD, NEW YORK, STANFORD ACHIEVLMENT TEST, LIPPINCOTT BASIC READING SERIES, EARLY TO READ ITA PROGRAM, SCOTT FORESMAN NEW BASIC READERS SERIES

BASAL READER SYSTEMS FOR BEGINNERS WERE ANALYZED TO DETERMINE THE EFFECTS OF SPECIFIC SYSTEM FEATURES ON THE READING ACHIEVEMENT OF FIRST-GRADE CHILDREN OF DIFFERENT SEX AND LEVELS OF INTELLIGENCE. THE PROJECT COVERED THE FOLLOWING SYSTEMS--(1) THE LIPPINCOTT "BASIC READING" SERIES, (2) THE "EARLY-TO-READ INITIAL-TEACHING-ALPHABET" PROGRAM, AND (3) THE SCOTT-FORESMAN "NEW BASIC READERS" SERIES. APPROXIMATELY 650 CHILDREN IN 26 FIRST-GRADE CLASSES WERE INCLUDED IN THE EXPERIMENTAL SAMPLE. THE INTERACTIONAL EFFECTS OF THE THREE EXPERIMENTAL TREATMENTS, THE THREE LEVELS OF INTELLIGENCE, AND THE TWO SEXES WERE ANALYZED. THE LIPPINCOTT AND "EARLY-TO-READ" PROGRAMS UTILIZED AN APPROACH TO READING THAT IS CONSIDERABLY MORE ANALYTIC THAN THE SCOTT-FORESMAN PROGRAM. STUDENTS USING THE LIPPINCOTT SERIES EXPERIENCED SIGNIFICANTLY HIGHER ACHIEVEMENT ON THE SUBTESTS OF VOCABULARY AND SPELLING IN THE "STANFORD ACHIEVEMENT TEST" THAN THE OTHER TWO GROUPS. AMONG ANY OF THE THREE BASAL READER SYSTEMS, THE FACTOR OF SEX DID NOT PRODUCE DIFFERENTIAL RESULTS. INTELLIGENCE WAS NOT A MAJOR FACTOR IN DISTINGUISHING PROBABLE CHANGES FOR SUCCESS. ON THE VARIABLES OF POSTINSTRUCTION ORAL READING AND WORD RECOGNITION, THE LIPPINCOTT AND "EARLY-TO-READ" STUDENTS WERE COMPARABLE WHEN MEASURED. BOTH GROUPS WERE SIGNIFICANTLY SUPERIOR TO THE SCOTT-FORESMAN PUPILS ON THESE SPECIFIC VARIABLES. (JH)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

EFFECTIVENESS OF
THREE DIFFERENT BASAL READING SYSTEMS
ON FIRST-GRADE READING ACHIEVEMENT

Cooperative Research Project No. 2720

Harold J. Tanyzer
Harvey Alpert

Hofstra University
Hempstead, New York

1965

The research reported herein was supported by the
Cooperative Research Program of the Office of Education,
U.S. Department of Health, Education and Welfare

ED003485

ACKNOWLEDGMENTS

Sincere appreciation is expressed to the three supervisors -- Mr. William A. Bacci, Miss Mary Miller, and Miss Dorothy Dietrich -- and to the twenty-six first-grade teachers of the East Meadow, Carle Place, and Uniondale school districts who participated in this study. The investigators are also grateful to the district superintendents -- Mr. John H. Moehle of Carle Place, Dr. Edward J. McCleary of East Meadow, and Dr. John J. Forester of Uniondale -- and to their administrative staffs for their cooperation and support.

Special thanks are expressed to the Hofstra Research Staff composed of Mrs. Lenore Sandel, whose invaluable suggestions and care of records made a complex task bearable; Mrs. Lois Young, and Mrs. Tess Hyams, who observed in the classrooms to verify experimental conditions; and to Mrs. Roslyn Newman, for her calm, cheerful acceptance of deadlines and the frequent revisions in the typing of the manuscript. We also acknowledge our indebtedness to Dr. Leslie D. McLean, our statistical consultant, for his suggestions and aid in the treatment of the data. Without his assistance, this study would not have been possible.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iii

LIST OF TABLES vii

LIST OF FIGURES xi

CHAPTER

I. STATEMENT AND DEFINITION OF THE PROBLEM 1

Statement of the Problem
Objectives and Hypotheses
Related Research

II. PROCEDURE 9

Description of the Basal Reader Systems
Selection of the Sample
Teacher Selection and Training
Supervision of the Program
Instructional Control
Meetings with Teachers and Research Officers
Evaluation Measures
Treatment of the Data

III. RESULTS AND DISCUSSION OF FINDINGS 29

Description of the Sample: Community Information
Description of the Sample: Teacher Information
Description of the Sample: Pupil Information
Results of the Reading Readiness and Intelligence Tests
Results of the Teachers' Logs
Evaluation of Teacher Effectiveness
Results of Post-instruction Tests on Subsamples
Results and Discussion: Reading Achievement
Use of Readiness Tests to Predict Reading Achievement

IV. SUMMARY AND CONCLUSIONS 104

APPENDIX

A. Teacher's Log of Reading Instruction and Supportive Activities 111

B. Teacher Rating Scale and Criteria 114

APPENDIX Page

C. Word Recognition Tests Administered to Random Sample: Gates Word Pronunciation Test, Karlsen Phonemic Word Test, Phonetically Regular Words Oral Reading Test 116

D. Correlation Matrix of the Reading Readiness Tests and the Stanford Word Reading, Paragraph Meaning, and Word Study Skills Subtests for the Lippincott Treatment Group 121

E. Correlation Matrix of the Reading Readiness Tests, and the Stanford Word Reading, Paragraph Meaning, and Word Study Skills Subtests for the i/t/a Treatment Group 122

F. Correlation Matrix of the Reading Readiness Tests and the Stanford Word Reading, Paragraph Meaning, and Word Study Skills Subtests for the Scott, Foresman Treatment Group 123

G. Partial Correlations between Predictors: Murphy-Durrell Phonemes, Murphy-Durrell Letter Names, Metropolitan Total Score, and Reading Achievement Criteria: Stanford Word Reading, Paragraph Meaning, and Word Study Skills for Lippincott Boys, Girls, and Total Group 124

H. Partial Correlations between Predictors: Murphy-Durrell Phonemes, Murphy-Durrell Letter Names, Metropolitan Total Score, and Reading Achievement Criteria: Stanford Word Reading, Paragraph Meaning, and Word Study Skills for i/t/a Boys, Girls, and Total Group 125

I. Partial Correlations between Predictors: Murphy-Durrell Phonemes, Murphy-Durrell Letter Names, Metropolitan Total Score, and Reading Achievement Criteria: Stanford Word Reading, Paragraph Meaning, and Word Study Skills for Scott, Foresman Boys, Girls, and Total Group 126

REFERENCES 127

PRECEDING PAGE BLANK-NOT FILMED

LIST OF TABLES

Table		Page
1.	School District Data 1963-64 School Year	16
2.	Distribution of Pupils in the Sample in September 1964	17
3.	Analysis of Variance Design	28
4.	Number of First-Grade Classrooms in Each District	30
5.	Highest Degree Held by Teachers	34
6.	Type of Teaching Certificate Held by Teachers	35
7.	Marital Status of Teachers	36
8.	Number of Married Teachers' Own Children	37
9.	Distribution of Pupils by Sex in the Three Treatment Groups in September 1964	38
10.	Distribution of Pupils by Ethnic Class in the Three Treatment Groups	40
11.	Pre-First-Grade School Experience of the Pupils in the Three Treatment Groups in September 1964	41
12.	Pre-Instruction Reading Ability of Selected Pupils in each Treatment Group as Measured by the Detroit *Word Recognition Test*	42
13.	Means and Standard Deviations of IQ Scores as Measured by the Pintner-Cunningham Primary Test for the Three Treatment Groups	43
14.	Means and Standard Deviations of Reading Readiness Scores of the Three Treatment Groups	46
15.	Significance of the Differences between Treatments in Average Number of Minutes Spent per Day by Teachers in Direct Reading Instruction and Supportive Activities, Fall 1964	53
16.	Significance of the Differences between Treatments in Average Number of Minutes Spent per Day by Teachers in Direct Reading Instruction and Supportive Activities, Spring 1965	54

Table Page

17. Number of Reading Groups by Percentage of Teachers within Each Treatment Fall and Spring 1964-65 56

18. Percentage of Pupils at Each Instructional Reader Level of the Scott, Foresman Population 58

19. Percentage of Pupils at Each Instructional Reader Level of the Lippincott Population 59

20. Percentage of Pupils at Each Instructional Reader Level of the i/t/a Population 60

21. Percentage of Teachers in Each Treatment Group as Rated on a Five Point Scale in Class Structure, Class Participation, Attention to Individual Needs, and Over-all Teacher Competence 63

22. Significance of the Differences between Mean Grade Equivalence Scores in Accuracy on the Gilmore Oral Reading Test of Randomly Selected Pupils within the Three Treatments 67

23. Significance of the Differences between Mean Number of Words Read per Minute (Rate) on the Gilmore Oral Reading Test of Randomly Selected Pupils within the Three Treatments 68

24. Significance of the Differences between Mean Scores on the Gates Word Pronunciation Test of Randomly Selected Pupils within the Three Treatments 69

25. Significance of the Differences between Mean Scores on the Karlsen Phonemic Word Test of Randomly Selected Pupils within the Three Treatments 70

26. Significance of the Differences between Mean Scores on Phonetically Regular Words Oral Reading Test for Randomly Selected Pupils within the Three Treatments 71

27. Significance of the Differences between Mean IQ Scores on the Pintner-Cunningham Primary Test of Randomly Selected Pupils within the Three Treatments 72

28. Number and Mean IQ Scores of Subjects Classified by IQ Level, Sex, and Treatment 75

29. Table of Means for the Three Treatment Groups on the Composite Reading Score (Word Reading, Paragraph Meaning, Word Study Skills of the Stanford Achievement Test) 76

Table Page

30. Analysis of Variance Table for Composite Reading Score 80

31. Table of Means for the Three Treatment Groups on Word Reading Subtest of the Stanford Achievement Test 81

32. Analysis of Variance Table for Word Reading Subtest of the Stanford Achievement Test 83

33. Table of Means for the Three Treatment Groups on Paragraph Meaning Subtest of the Stanford Achievement Test 85

34. Analysis of Variance Table for Paragraph Meaning Subtest of the Stanford Achievement Test 86

35. Table of Means for the Three Treatment Groups on Word Study Skills Subtest of the Stanford Achievement Test 89

36. Analysis of Variance Table for Word Study Skills Subtest of the Stanford Achievement Test 90

37. Table of Means for the Three Treatment Groups on Vocabulary Subtest of the Stanford Achievement Test 92

38. Analysis of Variance Table for Vocabulary Subtest of the Stanford Achievement Test 94

39. Table of Means for the Three Treatment Groups on Spelling Subtest of the Stanford Achievement Test 95

40. Analysis of Variance Table for Spelling Subtest of the Stanford Achievement Test 96

41. Canonical Correlations by Sex and Treatment 100
Predictors: Murphy-Durrell Phonemes, Murphy-Durrell Letter Names, Metropolitan Total Score, Pintner-Cunningham IQ
Reading Achievement Criteria: Stanford Achievement Word Reading, Word Study Skills, and Paragraph Meaning

PRECEDING PAGE BLANK-NOT FILMED

LIST OF FIGURES

Figure **Page**

1. Frequency Distributions of IQ Scores on the Pintner-Cunningham Primary Test of the Three Treatment Groups 44
2. Bivariate Distribution of Composite Reading Scores and IQ Scores of the Lippincott Population 77
3. Bivariate Distribution of Composite Reading Scores and IQ Scores of the i/t/a Population 78
4. Bivariate Distribution of Composite Reading Scores and IQ Scores of the Scott, Foresman Population 79
5. Relationship between Treatment and IQ on Stanford Word Reading 84
6. Relationship between Treatment and IQ on Stanford Paragraph Meaning 88
7. Relationship between Treatment and IQ on Stanford Word Study Skills 91
8. Relationship between Treatment and IQ on Stanford Spelling 97

CHAPTER 1

STATEMENT AND DEFINITION OF THE PROBLEM

It is generally recognized that basal readers provide the framework for reading instruction in the majority of first-grade classes in American schools. Despite certain criticisms of basal readers, they have maintained their popularity and gained widespread acceptance. In recent years, however, alternative programs for teaching reading have appeared that exhibit discrete features differing from the conventional basal reader series. Since, for the most part, children tend to learn to read in response to the content and structure of the teaching materials, it seems reasonable to expect that different programs of reading instruction are likely to lead to different reading performances among children of varying intelligence and readiness for reading.

Statement of the Problem

The major problem on which this research focused was the evaluation of the effectiveness of three different basal reader systems on the reading *achievement of first-grade children differing in intelligence and sex.* These programs were (a) the Lippincott series, an intensive phonics system by McCracken and Walcutt, (b) a series by Mazurkiewicz and Tanyzer, which utilizes a special beginner's teaching medium known as the Initial Teaching Alphabet (i/t/a), and (c) the Scott, Foresman series by Robinson, Monroe, and Artley, which is eclectic in that it combines many approved methods of teaching reading and is typical of the most frequently used basal reader programs in the United States.

It is possible that each of the three systems under study produces differentiated results in terms of the intelligence of the child and the various maturational factors measured in readiness tests. Hence, a subsidiary purpose of this study was whether it was possible to classify children on the basis of intelligence and readiness factors to predict the reader system which would produce the best results for the individual. Since no one method is apt to prove best for all children, it was hoped that this study would help to reveal which cluster of readiness factors results in the most favorable prognosis in which one of the three systems, resulting in a greater differentiation of approaches in teaching beginning reading than currently exists in this country.

Children of different levels of intelligence learn differently, and it is known from previous research that intelligence is highly related to reading. Hence, there is great concern and little factual knowledge with respect to the effect of a system upon the reading attainments of children of varying levels of mental ability. The most commonly used systems attempt to compensate for these differences by using an eclectic approach which combines a variety of methods without a strong emphasis upon any one procedure. Systems developed in which methodological emphases do exist may be of greater benefit to children of high than low intelligence or vice versa.

A difference in reading achievement in first-grade has also been observed and reported in research studies indicating that there is a higher incidence of reading difficulties among males than females. Thus, it was important to determine whether any one of the three systems produces greater success with males or females. If such evidence was obtained, then a subsequent improvement in differentiating instruction on the basis of sex might prove to lessen or eliminate the gap in reading achievement among boys and girls of approximately similar levels of intelligence.

Objectives and Hypotheses

The following null hypotheses were tested with respect to first-grade children taught by the three basal reader systems:

Hypothesis 1. There are no significant differences on an achievement test of reading and spelling (word reading, paragraph meaning, vocabulary, spelling, and word study skills) at the end of the first-grade.

Hypothesis 2. There are no significant differences on an achievement test of reading and spelling (word reading, paragraph meaning, vocabulary, spelling, and word study skills) among children of different intelligence levels (high: upper third; average: middle third; and low: bottom third).

Hypothesis 3. There are no significant differences on an achievement test of reading and spelling (word reading, paragraph meaning, vocabulary, spelling, and word study skills) between males and females.

Hypothesis 4. There are no significant differences on oral reading and word recognition tests among small subsamples selected randomly from each basal reader group.

Related Research

i/t/a Studies

Sir James Pitman (1961) developed the Initial Teaching Alphabet as a simplified alphabet for improving the learning of reading during the beginning stages of instruction. His intention was that the child should use i/t/a as a temporary transitional medium until he developed fluency and a reasonable degree of competence in reading in i/t/a, and then should transfer to reading books printed in traditional orthography (T.O.). i/t/a was designed to overcome the visual and auditory inconsistencies of English orthography by providing a more reliable printed code for the phonemes of English. Each of i/t/a's 44 characters represents one, and only one, basic sound unit. Twenty-four of its characters are retained from the conventional alphabet, and fourteen others resemble two familiar letters joined together. The medium of i/t/a and its spelling rules were deliberately designed to resemble traditional orthography, in order that the transition from i/t/a to T.O. would be a relatively easy adjustment.

Research studies (Downing, 1964b; Mazurkiewicz, 1964b) have proven that children can learn to read more easily in i/t/a than in T.O., and that the reading skill they are able to achieve in i/t/a is superior to that attained by children instructed in T.O. initially.

Relatively little evidence is reported on the post-transition reading performance of children taught with i/t/a as to whether they can read traditional orthography better than children who are taught in traditional orthography initially. Downing (1964a) reported on the transfer results for a subsample of 146 pupils who were between the ages of five and one-half and six and three-quarters and had been in school for slightly over one and one-half years at the time they were tested on the standard, or T.O. edition of the Neale Analysis of Reading Ability. He found that the i/t/a subsample obtained superior comprehension and accuracy scores than that attained by a matched control group who had been taught in T.O. from the beginning of their schooling. It was also noted that more than one-half of the i/t/a pupils, at the time they were tested, had not been transferred to T.O. books by their teachers.

Mazurkiewicz (1964a) reported finding no significant difference between first-grade children taught in i/t/a and those instructed in traditional orthography when both groups were tested in May on the standard edition of the California Reading Test. This finding differed from Downing's results which indicated a significant difference between the experimental and control groups. This would seem to suggest that time might be an important variable in discovering differences between children taught to read in i/t/a as compared to T.O. Mazurkiewicz (1964a) noted that, at the time of testing, only 26% of the i/t/a group had been transferred to instruction in T.O., and that a larger percentage of culturally deprived children were included in the i/t/a group. When the i/t/a pupils who had made the transition were matched with pupils in the control population, the i/t/a group's performance on the vocabulary subtest of the California Reading Test was significantly higher. It is possible that a sampling bias existed in favor of the i/t/a group, since it included only those pupils who had made transition, suggesting that they were among the best readers in the i/t/a first-grade population.

Relatively little evidence is available concerning the effect of i/t/a on the spelling ability of the child. Incomplete results reported by Downing (1964b) indicate that by the middle of the third year of their schooling, i/t/a pupils were superior to T.O. pupils in spelling achievement on a test administered in traditional orthography.

Sex differences in beginning reading

Studies on sex differences in reading readiness and reading achievement (Carroll, 1948; Nila, 1953; Samuels, 1943; Wilson, Burke, & Fleming, 1939) have indicated that girls are superior to boys.

Prescott (1955), in a large-scale study of over 14,000 first-grade children, reported that with chronological age equated, the performance of girls on the Metropolitan Readiness Tests was significantly superior to that of boys.

Heilman (1961, p. 353), in his review of the literature in the area of sex differences, concluded that the balance of evidence suggested that girls exceeded boys in reading achievement in each of the primary grades, but that this superiority tended to diminish during the intermediate grades.

Generally then, more girls than boys are likely to achieve success in beginning reading, and boys are less apt to be ready for reading instruction when they enter school. Many reasons have been advanced for this phenomenon, but the one mentioned most frequently is immaturity in social, physical, and emotional areas (Harris, 1961, p. 27; Heilman, 1961, pp. 358-359).

Since i/t/a is a more consistent medium than T.O., less maturity in areas such as physical and motor control should be necessary for success in beginning reading. Hence, in this study, sex was included as an independent variable, and separate measurements were made for boys and girls. If i/t/a or more analytic procedures to the teaching of reading, such as is true in the Lippincott series, are primary features of the methods or of the application of these methods, then the results with boys should approximate those obtained with girls in either the i/t/a or the Lippincott procedure more than in the Scott, Foresman program.

Relationship between reading and intelligence

General intelligence, expressed in terms of mental age or intelligence quotient, is acknowledged as being highly related to success in beginning reading, with mental age giving the better prediction of success (Harris, 1961, p. 28; Wilson & Fleming, 1937).

There is considerable evidence suggesting that there are differences in the way children of high, average, and low intelligence learn. Torgerson and Adams (1954, pp. 94-95), in discussing the characteristics of dull and bright children and the implications for teaching each type, suggest that children of high intelligence generally are able to grasp generalizations and apply these generalizations in new situations without difficulty. Children of average intelligence can learn generalizations, but require some specific teaching in developing the proper application of generalizations. With children of low intelligence, generalizations have been found not to be useful; although these children are capable of memorizing generalizations, they are not able to make the associations necessary to know when these generalizations should be applied, without specifically being told.

One of the goals of education is to differentiate instruction sufficiently so that each child will be taught by procedures in which the probability of success will be maximized. Hence, the children who were taught in each of the three basal reader systems in this study were divided not only by sex, but in terms of intelligence -- using three categories, high, average, and low. By doing this, it would be possible to determine whether any of the methods have a differential effect, prove more successful with children of high, average, or low intelligence than one of the other procedures.

Predictive value of readiness tests

Reading readiness tests have been found to be somewhat predictive of success in beginning reading (Bremer, 1959; Karlin, 1956; Lee, Clark, & Lee 1934). However, the relationship between reading readiness scores and reading achievement is somewhat lower than the relationship between intelligence scores and reading (Dean, 1939; Deputy, 1930).

Readiness tests were employed in the study herein to determine whether any differential effect would result in the prediction of beginning reading in any of the three systems for males and females of high, average, and low intelligence. This would answer two questions: (a) whether the three basal reader systems are, in fact, different in terms of whether different readiness factors produce better predictions of success for certain kinds of children in one system than another, and (b) whether it is possible to differentiate among children to the point where we can determine which system can provide or predict the best likelihood of success.

CHAPTER 2

PROCEDURE

The major purpose of this study was to evaluate the effectiveness of three different basal reader systems, each containing discrete features, on the reading attainments of first-grade children of different levels of intelligence and sex. In connection with this problem, it was hypothesized that there would be no significant differences among the three treatments in reading achievement. This chapter is concerned with the procedures used to gather the necessary data to test the hypotheses which were stated in Chapter 1. First, however, the three basal reader systems evaluated in this study are described.

Description of the Basal Reader Systems

Basic Reading[1] (This independent treatment variable is hereafter referred to as Lippincott.)

The authors (McCracken and Walcutt, 1963, p. iv.) define reading, in the teacher's edition, as follows: "Reading is first of all, and essentially, the mechanical skill of decoding, of turning the printed symbols into the sounds which are language." This series has a single pre-primer, one primer, and two first readers, a workbook for each reader, and manuals for teachers. There are no separate readiness materials, although reference is made to "functional readiness" activities which are included in the eleven lesson plans of the pre-primer, but discontinued thereafter.

---

[1] Glenn McCracken and Charles C. Walcutt, Basic Reading. Philadelphia: J. B. Lippincott Company, 1963.

The methodology emphasizes early and intensive phonics. Letter sounds are taught by relating their sounds in spoken words to their corresponding spellings in printed word forms. Words are generally introduced only after their phonic elements have been taught. The teacher's manual states that more than 2,000 words are introduced in the first-grade program. A number of these words appear in lists removed from context.

Early-to-Read i/t/a Series[1] (This independent treatment variable is hereafter referred to as i/t/a or Early-to-Read i/t/a.)

The unique feature of this basal reader program is its use of Pitman's Initial Teaching Alphabet. The program is designed for use in first-grade, and consists of a set of readers and workbooks, including a readiness workbook, six readers, and four workbooks which accompany the readers. Also included are alphabet and word cards and an alphabet book for use with the texts. Teachers' manuals accompany the texts and workbooks, and an i/t/a Handbook for Writing and Spelling is provided for the teacher.

The series stresses a language-experience approach together with the teaching of the forty-four characters of Pitman's alphabet. Pupils are taught to recognize and write each character of the new alphabet and to associate its visual form with its speech equivalent. Intensive training in visual and auditory discrimination accompanies the teaching of each character. After several characters have been introduced, pupils are taught to use an analytical procedure to unlock the pronunciation of new words. This approach emphasizes the analysis and synthesis of the characters of the word as the principle method of word identification. By the end of Book 3, all forty-four characters have been presented. In Books 4, 5, and 6, structural-analysis skills are introduced, including compound words, a number of inflectional endings, such as contractions and possessives, and other suffixes which show a change of number, tense, or comparison. Syllabication skills are briefly introduced in Book 6. Book 7, the last book in the series, is the "transition" reader. In the first part of this book, pupils are taught to recognize the most common graphemes in traditional orthography for each of the i/t/a characters. Conventional spelling(s) for each of the phonemes of English are gradually introduced and substituted for each of the i/t/a characters. The last half of Book 7 is printed entirely in traditional orthography.

---

[1] Albert J. Mazurkiewicz and Harold J. Tanyzer, Early-to-Read i/t/a Series. New York: Initial Teaching Alphabet Publications, 1963.

The New Basic Readers[1] (This independent treatment variable is hereafter referred to as Scott, Foresman.)

The materials for the first-grade program include a set of texts, consisting of a core reading readiness book and two other readiness books for slow learning or immature pupils, three pre-primers, a junior primer, a primer, and a first reader. Also included are workbooks correlated with the readers, a set of manuals for teachers for both the texts and the workbooks, and supplementary teaching aids such as vocabulary and phrase cards, an introductory "big book" and a picture dictionary.

Following readiness instruction, the new words appearing in the three pre-primers are taught as sight words through a "look-say" or whole word approach. Gradually, this method gives way to the development of interrelated word analysis skills, including phonics, structural-analysis, and use of contextual clues. Phonic elements, beginning with initial consonants, are developed from sight words and taught through an inductive approach. The procedure for teaching a phoneme involves hearing the sound within spoken words, identifying the printed symbol for that sound in several known words containing it, and associating the symbol with the sound it represents. The "substitution technique" is taught to pupils as an important aid in identifying a new word that resembles a known one, except for an initial or final element.

---

[1] Helen M. Robinson, Marion Monroe, and A. Sterl Artley, The New Basic Readers. New York: Scott, Foresman and Company, 1962.

Vocabulary is introduced at a carefully controlled rate, and each new word is repeated at least ten times in the book in which it is first introduced and then repeated a minimum of ten more times in each of the subsequent readers. Altogether, 323 new words are introduced.

The following section summarizes and compares the discrete features of the three basal reader systems used in this study.

Printed medium

Lippincott and Scott, Foresman use conventional orthography as the instructional medium, while the i/t/a series uses Pitman's Initial Teaching Alphabet.

Vocabulary load

The vocabulary count is considerably higher in the i/t/a and Lippincott series than in the Scott, Foresman series. Scott, Foresman introduces a total of 323 words in its first-grade program, while Lippincott reports a total of over 2000 words, and i/t/a introduces approximately 1200 words in its readers. All the words appearing in the

Scott, Foresman and i/t/a readers are woven into the printed context, while Lippincott introduces a number of words in lists. Each list contains words illustrating a specific phoneme-grapheme relationship.

Frequency of word repetition

There is less control over the introduction of new words and significantly fewer word repetitions in the i/t/a and Lippincott series as compared to the Scott, Foresman program with its carefully controlled vocabulary and high frequency of repetitions.

Content of the first-grade phonic program

Of the three series, Lippincott has the most intensive phonic program. The program introduces over 100 letters and letter combinations including long and short vowels, vowel digraphs, dipthongs, silent vowels, single consonants, consonant blends and digraphs. Also introduced in the program are a number of phonic generalizations and spelling rules. Beginning with pre-primer reading, the short vowels are taught initially. The i/t/a series also emphasizes an early introduction of phonics. The forty-four characters of Pitman's alphabet are taught in the first two workbooks and readers of the series. The lightest phonic program is carried by the Scott, Foresman series which introduces all the single consonants and a number of two-letter consonant combinations. No vowels are introduced in the first-grade program.

Content of the first-grade structural-analysis program

The Lippincott program teaches a number of inflectional endings, compound words, and introduces syllabication. The Scott, Foresman series introduces a number of inflected, compound, and derived forms. The i/t/a series developes the largest number of structural-analysis skills and introduces syllabication after children have learned the forty-four characters of Pitman's alphabet. Children are taught a large number of inflectional endings and compound words.

Procedure for teaching word recognition

In the Scott, Foresman program, a sight vocabulary is developed initially before any phonics is introduced. Beginning with pre-primer reading, each new word is taught by procedures which emphasize recognizing the word as a whole. A phonic element is taught to children after several words containing it have been learned as sight words. Children are helped to discover for themselves the relationship between a letter and its corresponding sound value. Word analysis is gradually developed through an eclectic approach which stresses a combination of phonic, structural, and context clues.

The Lippincott series does not develop an initial sight vocabulary, instead, it teaches letter sounds by relating individual sounds in spoken words to their corresponding spellings in printed words. The initial vocabulary consists primarily of phonetically regular words which are selected on the basis of the phonic elements they contain. New words are not introduced until their component letter sounds have been taught. Some phonetically irregular words are taught as they are used in the stories. Their number is held to a minimum.

The i/t/a series emphasizes a combined experience story and phonics approach. Children are taught to associate each of the forty-four characters of Pitman's alphabet with the phoneme it represents. The primary method of identifying an unfamiliar word involves analyzing and synthesizing the sound elements of the word. A number of new words appearing in the beginning readers and experience stories are taught as sight words, if they contain characters not previously introduced. The order of presentation of the i/t/a characters is primarily based on a study of the frequency of usage of a sound in the speech and writing of children (Dewey, 1923). Writing activities which stress the phoneme-grapheme relationships are incorporated in the word recognition program. Beginning in Book 6 and continuing through the first half of Book 7, each of the i/t/a characters is gradually replaced, in words containing it, by its corresponding T.O. spelling.

Since the main thrust of this study was to determine whether any differential effects occur in first-grade reading achievement with respect to the use of three different basal reader systems, the discrete features of each series were described and the major differences among the series were compared in the preceding section.

Selection of the Sample

Three Long Island school districts in Nassau County, New York -- East Meadow, Carle Place, and Uniondale -- were selected for this study on the basis of promised cooperation with the experimental conditions of the study and comparability with respect to mean family income, entrance age to grade 1, mean intelligence of first-grade children, and reading achievement at the end of grade 1 for the 1963-64 school year. Table 1 illustrates the information provided by each school district.

TABLE 1

SCHOOL DISTRICT DATA
1963-64 SCHOOL YEAR

School System	Intelligence Level of First-Grade Population	Mean Family Income	Entrance Age to Grade 1	Reading Achievement at End of Grade 1
East Meadow	Range: 75-140 Mean : 109	$8100	Must be 6 before 12/5	2.1
Carle Place	Range: 75-143 Mean : 113	8400	Must be 6 before 12/5	2.1
Uniondale	Range: 75-144 Mean : 109	8200	Must be 6 before 12/5	2.1

In general, the three school systems draw from relatively equivalent socio-economic levels. All children enter school at the kindergarten level and receive no formal reading instruction until first-grade. Nine first-grade classes in East Meadow and nine classes in Uniondale were selected randomly. Carle Place contributed its entire first-grade population of eight classes. Table 2 illustrates the distribution of pupils in the original sample.

As can be seen in Table 2, (on the following page), 701 pupils in twenty-six classes were in the original sample. However, there was some attrition over the experimental period of 140 days for various reasons, including such factors as incomplete data caused by a child being absent on one of the testing dates or moving from the community. The final sample approximated 650. The original populations of East Meadow and Uniondale were virtually identical with respect to the number of pupils and classes

TABLE 2

DISTRIBUTION OF PUPILS IN SAMPLE
SEPTEMBER 1964

District	Number of Pupils	Number of Classes	Average Class Size
Carle Place	206	8	25.8
East Meadow	248	9	27.8
Uniondale	247	9	27.4

and average class size, whereas the Carle Place sample had approximately forty fewer pupils, one less class, and a slightly smaller mean class size than either of the other two districts. However, these differences were not considered major factors in this study.

A major factor influencing the selection of samples from Uniondale, Carle Place, and East Meadow was that each one of these districts had used or planned to use one of the basal series that would be a variable in this study. Carle Place had been using the Lippincott series and the latest edition of the Scott, Foresman series was being employed in the Uniondale schools. The East Meadow school district had planned to experiment with the new i/t/a series, independent of this study. Unfortunately, it was not possible to obtain school districts willing to experiment with all three basal series. If this had been possible, it would have permitted random assignment of teachers to variables and of variables to schools, a preferred procedure in research.

Thus, the basal reader system used with each one of the three samples was as follows:

Carle Place: Basic Reading, published by J. B. Lippincott Company, 1963.

East Meadow: Early-to-Read i/t/a, published by i/t/a Publications, 1963.

Uniondale: The New Basic Readers, published by Scott, Foresman and Company, 1962.

Teacher Selection and Training

Each school district was requested to recruit volunteers for participation in the study. From among the volunteers, selections of the participating teachers were made at random, except in the Carle Place school district where all eight first-grade teachers in the system were selected. Prior to the opening of school, in September, the teachers in each of the three treatment groups attended a three-day workshop where training by variable was given to keep the methodologies of the three basal series distinct. Each workshop was conducted by a reading consultant or author representing the publisher of the basal series. Teachers were instructed in methods and practices of teaching reading as recommended by the authors of the series which they would use. Emphasis was given to adapting instruction to the learning needs of children varying in levels of ability and rates of learning. Methods of differentiating instruction were presented to the teachers in each program. Each of the following proposals was given consideration:

Differences in levels of ability

Basal readers and instructional materials (commercial and teacher-made) would be adapted to the abilities of the children.

Differences in rates of progress

Children would progress at different rates with respect to skills development. Flexible grouping procedures would be emphasized with groups changing in membership as individual progress warranted the change. Children may work in groups, in pairs, or by themselves, depending on the purpose of the activity and the learning needs of the individual or group.

Differences in skills needs

Emphasis was placed on evaluating pupil progress and diagnosing reading difficulties. Teachers were encouraged to prepare intensive practice materials, when feasible, in order to correct specific deficiencies, and to form groups on the basis of pupils' specific needs or performance in various reading skills.

Individualizing reading

The better readers would be encouraged to do more independent recreational reading in easy and interesting materials. Provision for individual teacher-pupil conferences would be made when it was appropriate, as in evaluating a child's reading, giving help in a particular skill, and so forth.

In each workshop, consideration was also given to enriching the instructional program. A wide variety of library books would be made available for pupil use. These books would cover a range of interest and reading levels and would be changed periodically during the school year. In the i/t/a classes, library books printed in the i/t/a medium would be provided. A number of juvenile T.O. books would be transliterated into i/t/a by the research staff. During the school year, as i/t/a pupils begin to transfer from i/t/a to T.O., library books printed in T.O. would be made available. Teachers were asked to set aside time in the reading program when pupils would have the opportunity to select books from the library corner for independent reading.

Whole-class activities would be carried on when they seemed appropriate, as in introducing new concepts, word meanings, or reading skills. In addition, the following types of related activities and learning aids would be used with the whole class: (1) listening to stories and poems read and told by the teacher; (2) group dramatization of a story; (3) use of the flannel board for acting-out a story; and (4) use of recordings, filmstrips, and motion pictures to develop new concepts, learnings, and understandings.

Teachers were encouraged to include interesting and varied reading material in the classroom reading center consisting of such materials as supplementary readers, easy books on social studies, science and health, teacher-prepared word games, puzzles, and reading charts.

Teachers were also encouraged to provide a variety of learning activities which support instruction in reading, such as spelling, handwriting, oral and written expression, and reading in content areas. It was recognized that variations in teaching practices were likely to occur with respect to these activities. Writing and spelling were to be taught in accordance with the procedures recommended in the teachers' manuals of each basal reader series and suggestions of the workshop instructors.

Supervision of the Program

In each school district, a research officer was appointed who was responsible for seeing to it that teachers followed the recommended instructional program and did not violate experimental conditions of the study. The research officer in each district was the elementary reading supervisor or specialist. The research officers were instructed to continue the supervisory activities which they normally assumed in their own districts. As such, they were directly responsible for supervising teachers and providing guidance with the particular basal reader program being used in the district.

Members of the research staff acted as observers and periodically visited classes in the participating districts on a biweekly basis. The purpose of these observations was to verify experimental conditions in the classrooms and to answer teachers' questions.

A policy was adopted whereby visitation to classes was confined to research officers and administrators of the school district and research personnel assigned to the project. The purpose of this policy was to maintain normal classroom conditions as much as possible and to minimize the effects, if any, of visitation by "outsiders." Each of the school districts involved in the project agreed to this policy and cooperated in maintaining and enforcing it.

Instructional Control

A daily time schedule for reading instruction was prepared that teachers of the three treatment groups were asked to follow. This time schedule dealt with the daily amount of time for reading instruction and supportive activities relating to reading. A teacher's log (see Appendix A) was developed by the research staff and distributed to the teachers. Each teacher was asked to record the daily program devoted to reading instruction and related learning activities. Since the teachers' logs were to be analyzed for time study, the design of the log provided for noting the activity, materials, and duration of time spent on each activity. A code was used to designate the various reading and supportive learning activities. Teachers were directed to devote approximately two and one-half hours per day in direct reading instruction and supportive activities. Direct reading instruction was defined as time spent by the teacher in direct supervision of learning activities in reading. Supportive activities were defined as time spent in teaching of reading other than the basal reader. An attempt was made to motivate all the teachers equally in the use of their basal reader system. The results of the teachers' reports, as determined by their logs, is presented in Chapter 3, under the heading *Results and Discussion of Findings*. Validation of the logs was determined by observation of the research staff who compared the logs they kept when visiting the classroom with the teachers' own reports.

Meetings with Teachers and Research Officers

Prior to the opening of school, a general meeting was held with all of the teachers of the three treatment groups for the purpose of describing the nature and purpose of the study. Also, it was made known that the study was one of twenty-seven coordinated research projects on first-grade reading supported by the Cooperative Research Program of the United States Office of Education. Over the experimental period of the study, three additional meetings were held with all teachers and supervisors of each group. Teachers were encouraged to discuss problems relating to their instructional program and to share suggestions for teaching reading, creative writing, and other activities which support the reading program.

Research officers met periodically with the project director to discuss topics relating to the instructional program. These sessions gave the research officers an opportunity to exchange ideas and evaluate teacher and pupil progress. The importance of adhering to the experimental conditions of the study, supervision of teachers, testing and difficulties arising in administering the program, were major topics that were discussed at these meetings.

Evaluation Measures

Reading readiness and intelligence tests were administered as pretests during the period from September 23 through October 8, prior to the commencement of the experimental period. Tests were administered by members of the research staff, research supervisors, members of the guidance departments and psychological departments of the participating schools, with the teacher of the class acting as a proctor. All tests were administered under strict adherence to standard conditions, and no deviation from standard directions was allowed. Timing, on those tests requiring it, was done with a stop watch or interval timer, and limits were adhered to exactly. The pretests administered were:

1. Murphy-Durrell Diagnostic Reading Readiness Test, Revised, 1964
   - (a) Phonemes
   - (b) Letter Names Test: Capital Letter Names, Lower-case Letter Names
   - (c) Learning Rate Test
2. Thurstone Tests
   - (a) Pattern Copying
   - (b) Identical Forms

3. Metropolitan Reading Readiness Tests, Form A, Revised, 1966[1]
    - (a) Word Meaning
    - (b) Listening
    - (c) Matching
    - (d) Alphabet
    - (e) Numbers
    - (f) Copying
4. Pintner-Cunningham Primary Test, Form A, Revised, 1964
5. Detroit Word Recognition Test[2]

All post-instruction measures administered to the total sample at the end of 140 days of instruction were given in traditional orthography. The measures administered were the:

1. Stanford Achievement Test, Primary 1 Battery, Form X, 1963[3]
    - (a) Word Reading
    - (b) Paragraph Meaning
    - (c) Word Study Skills
    - (d) Vocabulary
    - (e) Spelling

The following dependent variable measures were administered to randomly selected subsamples within each of the treatment groups:

1. Gilmore Oral Reading Test, Form A
    - (a) Accuracy
    - (b) Rate

---

[1] All tests were administered as requested by the publisher. However, only the Word Meaning and Listening tests and the total score were used in this study.

[2] Administered to those entering first-grade children who were identified by the teacher as possessing some measure of reading ability.

[3] Arithmetic subtest not included.

2. Karlsen Phonemic Word Test
3. Gates Word Pronunciation Test
4. Phonetically Regular Words Oral Reading Test

All tests administered in the subsample were printed in traditional orthography and were given by members of the research staff with strict adherence to standard conditions. One person was assigned to each group to administer all measures.

Scoring of all tests was done by a panel of trained people under the supervision of a research coordinator. All tests were scored independently by two people to insure accuracy. In cases where scores disagreed, the tests were completely rescored and checked again. Raw scores for all the above tests were then entered on IBM cards, and after punching, were doubly verified to insure accuracy.

A form was prepared on which teachers could list the number of children at each instructional level within their classroom as of May 15. This data will be presented in the appropriate section of this report. In addition, the number of children from i/t/a classes who had made transition, as determined by their having completed Book 7 of the i/t/a series and having begun instruction in traditional orthography books, will also be presented in that section.

Treatment of the Data

The primary purpose of the first phase of this study was to determine whether any differential effects occurred in reading achievement as a result of the use of the Early-to-Read i/t/a program, the Scott, Foresman series, or the Lippincott program. Since intelligence would be considered an important variable, comparisons among the three basal reader systems were accomplished for high, middle, and low IQ children. This was done by obtaining the total distribution of IQ's for the children in the three treatment groups and dividing it into thirds (high, middle, and low). The comparisons among the three basal systems were then made for children of high intelligence, average intelligence, and low intelligence. The response to be analyzed was the mean of all subjects in each class who fell into the appropriate IQ groups. The rationale for using this procedure is that all subjects have shared a common experience and, hence, do not qualify as individual independent observations. Hence, the class was used as the unit of observation rather than each individual child (Campbell & Stanley in Gage ed., 1963, p. 193).

The dependent variables in reading achievement were measured on the following subtests of the Stanford Achievement Test, Primary I Battery, Form X: (a) Word Reading, (b) Paragraph Meaning, and (c) Word Study Skills. To improve the reliability of the measurement, these reading scores were combined into a single score called the composite reading score. This was determined by dividing the individual scores on each child by the variances of the total subtests. The several reading subtests are all relatively short and, hence, would tend to be individually less reliable than a composite score in the same way that a shorter test is usually less reliable than a longer one. Hence, the two main reasons for combining the reading subtests were to gain power and reliability in some measure of reading ability. The formula for determining the composite reading score (CRS) follows:

$$CRS = \frac{t_1}{s.d._1} + \frac{t_2}{s.d._2} + \frac{t_3}{s.d._3}$$

CRS = Composite Reading Score (computed for each subject).

t = individual score for each reading subtest of the *Stanford Achievement Test*.

1. *Word Reading*
2. *Paragraph Meaning*
3. *Word Study Skills*

After the composite reading score was obtained, the second step involved a regression analysis (bivariate distribution of IQ and CRS) to test for linearity. This analysis will be presented visually, under *Results and Discussion of the Findings*, as a scatter plot with the CRS on the *x*-axis and IQ score on the *y*-axis for the Lippincott, i/t/a, and Scott, Foresman groups. At this point, a 3x3x2 Analysis of Variance Design (Scheffe, 1959, p. 262) was used to analyze the interactional effects among the three basal reader systems, three levels of intelligence (high, average, and low) and two sexes. A separate analysis of variance was computed for the CRS and each of the following subtest scores of the *Stanford Achievement Test*: *Word Reading*, *Paragraph Meaning*, *Vocabulary*, *Word Study Skills*, and *Spelling*. The analysis of variance design is presented in Table 3.

Since the subsamples for the other outcome variable measures were extremely small and would be made even smaller if analyzed by sex and IQ, a simple *t*-test was computed among the mean scores on each of the subsample measures.

TABLE 3

ANALYSIS OF VARIANCE DESIGN

	Treatments	Lippincott N=		Initial Teaching Alphabet N=		Scott, Foresman N=	
	Sex	Boys	Girls	Boys	Girls	Boys	Girls
	High						
IQ	Average						
	Low						

CHAPTER 3

RESULTS AND DISCUSSION OF FINDINGS

Description of the Sample: Community Information

Generalizing the results of any study is always a risky proposition, since the results are meaningful in terms of the sample studied. If one is to determine the applicability of the results of a study for his own community, one must first determine whether the sample studied in the research approximates the sample in his own community. In order to accurately evaluate the data, it is important to understand the communities from which this sample is drawn and also the type of teachers employed in the district studied, since the Long Island area is not typical of the total school population of the United States. The sample of this study was drawn from three school districts -- East Meadow, from which the i/t/a population was drawn; Carle Place, from which the Lippincott sample was drawn; and Uniondale, from which the Scott, Foresman population was obtained. The preferred procedure in research would have been to use each of the three basal reading systems in each one of the school districts. Unfortunately, this was not practical as it was impossible to obtain school districts willing to attempt all three systems. Therefore, an attempt was made to select districts, which on the basis of experience and information provided, were fairly similar.

Length of the school day

In each of the three school districts, the length of the school day was exactly similar, five hours plus one additional hour for lunch.

Length of the school year

The length of the school year in each of the cooperating school districts was as follows: Carle Place and East Meadow had 180 school days and Uniondale had 182 days. This would not be a major variable in the study, however, since the time between the tests given before and after the training in each of the three districts was kept constant at 140 days.

Number of first-grade classrooms in district

The figures appearing in Table 4 give some idea of the comparative size of the three school districts studied.

TABLE 4

NUMBER OF FIRST-GRADE CLASSROOMS IN EACH DISTRICT

Carle Place	East Meadow	Uniondale
8	43	15

In Carle Place, the entire school district contained eight first-grade classrooms, which indicated that the entire first-grade population of the school district was included in the study. The Uniondale district had fifteen first-grade classrooms, reflecting the fact that this district was much larger than Carle Place. The East Meadow school district was the largest district in the sample, as can be seen from the fact that there were forty-three first-grade classrooms in the district.

Library facilities

The type of library facilities available to the class was similar for each of the three districts. Each of the districts had the services of a librarian available in the building, and each district encouraged teachers to make use of these services.

Median years of education[1]

The East Meadow and Carle Place school districts had a median of twelve years of education completed by the adults, while Uniondale was slightly lower: a median of eleven years of education completed by adults in the community.

Median income[1]

The median family income in the East Meadow and Carle Place communities were relatively similar, with median family incomes ranging between $8001-$9000. Uniondale reports a slightly lower median income than the other two communities: $7001-$8000.

Population[1]

The population of the communities for each of the three school districts was as follows: the East Meadow district was by far the largest, falling into the population range of 40,000-50,000; Uniondale had reported a population which fell into the 20,000-25,000 range, and Carle Place had, by far, the smallest population, falling into the 5,000-10,000 range. Each of these districts would be classified as unincorporated suburban, and each of the districts contains a population of over 2500 people.

---

[1] Derived from 1960 U.S. Census.

Description of the Sample: Teacher Information

In the East Meadow and Uniondale districts, teachers were asked to volunteer for participation in the study. From those teachers volunteering, selection was made by random numbers of those who would be included in the study. Teachers were informed prior to their volunteering that they would be expected to attend a workshop in which the recommended procedures for teaching in the basal reader system used in their school would be taught. The only exception to this was in Carle Place, where only eight first-grade classrooms existed. In that district no random selection of teachers was made, since all first-grade teachers were included in the study. Teachers selected from the three districts were aware, however, that this was to be an experiment, that they would be required to adhere to experimental conditions, and also attend a workshop describing procedures recommended for the basal reader system in which they would teach.

Age of teachers

The average age of the nine i/t/a teachers in East Meadow was 26 years. The average age of the eight Carle Place teachers using the Lippincott system was 32 years, and the average age of the nine teachers in Uniondale utilizing the Scott, Foresman system was 37 years. These figures suggest that considerable differences existed among the ages of teachers in each of the three districts. East Meadow had, by far, the largest number of young teachers, with ages ranging from 22 to 30. By contrast, Uniondale's participating teachers ranged in age from 22-56, by far the largest number of older teachers. Carle Place fell somewhere in between Uniondale and East Meadow in teachers' age, with ranges in age from 22-44.

Total number of years of teaching experience

From the previous data, the ages of the teachers in each of the three groups, one would infer a difference in maturity of the teachers in each group as well as differences in their previous experience. An examination of the differences in the total number of years of teaching experience of the teachers in each of the three districts verify this inference. Teachers using the i/t/a system had the smallest amount of previous teaching experience, averaging slightly over four years, while teachers in the Lippincott group averaged slightly over six years of previous teaching experience, and Scott, Foresman teachers averaged almost nine years of previous teaching experience. Although these data verify the inferences made from the ages of the teachers in the sample, the discrepancies in previous experience in teaching are not as great as one would have expected on the basis of discrepancy in teacher age.

Number of years of first-grade teaching experience

As might be expected from the previous data, the i/t/a teachers had the least amount of first-grade teaching experience, averaging slightly over one year. However, the Lippincott teachers and the Scott, Foresman teachers both averaged slightly over four years of previous first-grade experience, which suggests that the differences in ages in these two groups was not as significant a factor as it might have appeared.

Degrees held by teachers

Table 5 illustrates the degrees held by the teachers in each of the three treatment groups.

TABLE 5

HIGHEST DEGREE HELD BY TEACHERS

	Lippincott N=8	i/t/a N=9	Scott, Foresman N=9
Less than B.A.	0	0	1
More than B.A.	6	8	6
M.A.	1	0	1
More than M.A.	1	1	1

Although some differences in age and amount of experience were observed in Figures 4, 5, and 6, the figures appearing in Table 5 indicate that the professional training of the teachers was relatively similar. Only one teacher using the Scott, Foresman series had less than a bachelor's degree, six of the eight teachers in the Lippincott group had more than a bachelor's degree, eight of the nine teachers in the i/t/a group had more than a bachelor's degree, and six of the nine Scott, Foresman teachers had more than a bachelor's degree, which suggests that training among the teachers in each of the districts was relatively similar. One Lippincott teacher and one teacher in the Scott, Foresman group possessed a master's degree, while one teacher using each of the series possessed credits beyond the master's degree.

Certification

Table 6 represents the distribution of certificates held by teachers in each of the three treatment groups.

TABLE 6

TYPE OF TEACHING CERTIFICATE HELD BY TEACHERS

	Lippincott N=8	i/t/a N=9	Scott, Foresman N=9
Type of Certificate			
Standard	6	9	8
Higher	2	0	1

Table 6 indicates that every teacher in the study possessed the standard certificate required for teaching elementary school in New York State. Two of the Lippincott teachers in Carle Place and one Scott, Foresman teacher in the Uniondale district, in addition, had special certificates beyond the standard certificate necessary for New York State elementary teachers. These three certificates were primarily in the areas of administration and supervision.

Marital status

In examining the marital status of the teachers in each of the three treatment groups of the sample, the following information was obtained:

TABLE 7

MARITAL STATUS OF TEACHERS

	Lippincott N=8	i/t/a N=9	Scott, Foresman N=9
Single	4	7	2
Married	4	2	6
Widowed or Divorced	0	0	1

Table 7 reveals that of the eight teachers teaching in the Lippincott system in Carle Place, four were single, and four were married. Among the East Meadow teachers using the Early-to-Read i/t/a system, seven of the teachers were single and two were married. This might have been expected on the basis of the youth of the teachers in East Meadow. In the Uniondale district, two Scott, Foresman teachers were single, six were married, and one was widowed or divorced.

Married teachers' own children

Table 8 shows the number of the married teachers' own children.

TABLE 8

NUMBER OF MARRIED TEACHERS' OWN CHILDREN

	Lippincott N=4	i/t/a N=2	Scott, Foresman N=7
Number of Children			
0	2	1	3
1	1	0	2
2	1	1	1
3	0	0	1

Of the four married teachers using Lippincott, two teachers had no children, one had one child, and one had two children. Among the i/t/a teachers, only two of the nine teachers were married, one married teacher had no children of her own and one teacher had two children. Among the Scott, Foresman teachers, three of the married teachers were childless, two had one child, one had two children, and one teacher had three children.

Summary statement

The teacher is, obviously, an important variable in a study of this sort. Wherever possible, a system of random numbers was used in selecting teachers who had volunteered for this study. It is quite evident from the above description that considerable differences existed among the teachers of the three treatment groups in terms of age and previous experience with the exception of Carle Place, where random selection of teachers was not possible. It must be assumed that the existing differences occurred by chance. One possible cause for the observed differences may result from the fact that in Uniondale, where the Scott, Foresman system was used, there has been a lower turnover of teachers than in East Meadow, where i/t/a was employed, because of the salary structures of the communities. There is, however, sufficient variation among the three groups to suggest that differences which occurred in the achievement of children could have been due to the variations among the kinds of teachers of each of the three groups. The only area in which similarity was observed was in professional training of the teachers. Because of these variations, classrooms rather than individual students were used as the unit of observation in computing the analysis of variance. This would allow the investigators to observe whether the differences existing among the teachers in each of the three groups in terms of children's achievement was greater than the differences observed among the groups.

Description of the Sample: Pupil Information

Sex

Table 9 represents the breakdown by sex of the three populations in this study.

TABLE 9

DISTRIBUTION OF PUPILS BY SEX
IN THE THREE TREATMENT GROUPS IN SEPTEMBER 1964

	Lippincott N=206	i/t/a N=248	Scott, Foresman N=247
Boys	115 (55.8%)	131 (52.8%)	127 (51.4%)
Girls	91 (44.2%)	117 (47.2%)	120 (48.6%)

From Table 9 it can be seen that for the Carle Place group, learning to read in the Lippincott basal readers, there was a considerable imbalance between the number of boys and girls in the sample; 44.2% were girls and 55.8% were boys. Since the sample from this school district included the entire first-grade population, one would have to assume that this difference occurred by chance, even though statistically probability of this occurrence is less than one in one hundred. An imbalance between the number of boys and girls in the sample occurred, also, in the East Meadow district, where the children were instructed in the basal reader system utilizing i/t/a and the Uniondale district, where the children were instructed in the Scott, Foresman basal series. In the i/t/a sample, 52.8% were boys and 47.2% were girls. Although both of these samples were randomly selected, the sex differences resulting would be significant at the .05 level, and hence would occur by chance only five times in one hundred. The significance of these differences was eliminated as a factor in the final result by the utilization of sex as an independent variable. Hence, on the final data, the analysis of variance was computed on the dependent variables of reading and spelling of boys and girls separately.

Ethnic class

The populations of the three sample groups were relatively similar in ethnic class, as illustrated in Table 10 on the following page.

In the original population studying under the Lippincott system, 205 of the 206 pupils on whom data were available were white, and one child was of Puerto Rican extraction. There were no Mexican, Indian, Negro or Chinese children in this sample. Within the group studying in the Early-to-Read i/t/a series, again, the population was quite homogeneous

TABLE 10

DISTRIBUTION OF PUPILS BY ETHNIC CLASS
IN THE THREE TREATMENT GROUPS

	Lippincott N=206	i/t/a N=248	Scott, Foresman N=247
White	205	244	221
Negro	0	0	25
Puerto Rican	1	3	1
Mexican	0	1	0

-- 244 of the total 248 children were white, one child was of Mexican extraction, and three were of Puerto Rican parentage. This group, also, had no Indian, Chinese, or Negro children in the population. In Chicondale, where the Scott, Foresman basal reader system was used, the population, also, was almost completely white. In this group, 221 of the children fell in that category, with one child of Puerto Rican extraction, and 25 of the children were Negro.

Pre-first-grade school experience

Table 11 illustrates the pre-first-grade school experience of the children in each of the three treatment groups. Pre-first-grade experience included kindergarten and nursery school training. As can be seen from Table 11 (on the following page), the groups were relatively homogeneous. Almost all of the children in each of the three treatment groups received 101 to 200 half-days of school experience prior to entrance into first-grade. In the Lippincott and Scott, Foresman groups, only one child had

TABLE 11

PRE-FIRST-GRADE SCHOOL EXPERIENCE OF THE PUPILS IN THE THREE TREATMENT GROUPS IN SEPTEMBER 1964

	Lippincott N=206	i/t/a N=248	Scott, Foresman N=247
None	1	0	1
Less than 20 days	0	0	1
21-100 ½ days	0	0	0
101-200 ½ days	203	246	239
201-300 ½ days	2	2	4
301-400 ½ days	0	0	1
401-500 ½ days	0	0	0
501-600 ½ days	0	0	1

received no experience prior to his first-grade entrance, and in the Scott, Foresman group, one additional child had received less than 20 days of prior first-grade school experience. In the Scott, Foresman group, six children had received more than 200 half-days of previous school experience, while in the Lippincott and Early-to-Read i/t/a groups each had only two children who received more than 201 half-days of prior first-grade school experience. Hence, no significant differences from a practical standpoint existed in this area.

Pre-instructional reading ability

The Detroit Word Recognition Test was administered to those children entering first-grade who had been identified by the teacher as possessing some measure of reading ability prior to the start of the instructional program. The Detroit test consists of forty words and phrases, with matching pictures. The pupil reads a word or phrase, selects the picture to match it, and then draws a line from the word to the picture. Reading, as measured by the Detroit test, does not include comprehension but is primarily a measure of word recognition. The results of this measure, reported in raw scores, are illustrated in Table 12.

TABLE 12

PRE-INSTRUCTION READING ABILITY OF SELECTED PUPILS IN EACH TREATMENT GROUP AS MEASURED BY THE DETROIT WORD RECOGNITION TEST

	Lippincott N=38	i/t/a N=6	Scott, Foresman N=3
Mean Raw Score	9.8	10.3	27.3

As can be seen from Table 12, only six pupils from the i/t/a group, three from the Scott, Foresman group, and thirty-eight from the Lippincott population had been identified by their teachers as having some measure of reading ability. Statistical analysis of the mean scores was not possible due to the very small number of pupils to whom the Detroit test was administered. The large number of pupils tested in the Lippincott group, as compared to the very small number in the i/t/a and Scott, Foresman groups, would suggest the possibility that these differences in number resulted from differences in the kindergarten readiness programs which the children in the sample had experienced prior to first-grade entrance.

Results of the Reading Readiness and Intelligence Tests

Intelligence

Figure 1 represents the frequency distributions for the IQ scores, as measured by the Pintner-Cunningham Primary Test, for each of the three groups studied. From Figure 1, it can be seen that in the total groups some differences in intelligence did occur. Nevertheless, the distributions for each of the three groups was normal.

Table 10 represents the means and standard deviations for each of the three groups on IQ, as measured by the Pintner-Cunningham Primary Test.

TABLE 10

MEANS AND STANDARD DEVIATIONS OF IQ SCORES AS
MEASURED BY THE PINTNER-CUNNINGHAM PRIMARY TEST FOR
THE THREE TREATMENT GROUPS

	N	Mean	S.D.
Lippincott	208	103.6	13.4
i/t/a	248	102.3	13.3
Scott, Foresman	247	98.3	11.3

The mean IQ of the Lippincott group was 103.6, with a standard deviation of 13.4; for the i/t/a group, the mean IQ was 102.3 with a standard deviation of 13.3; and for the Scott, Foresman group, the mean IQ was 98.3 with a standard deviation of 11.3. The standard deviations were virtually identical. Critical ratios were computed between each of the possible pairs of mean IQ scores on the Pintner-Cunningham Primary Test. These

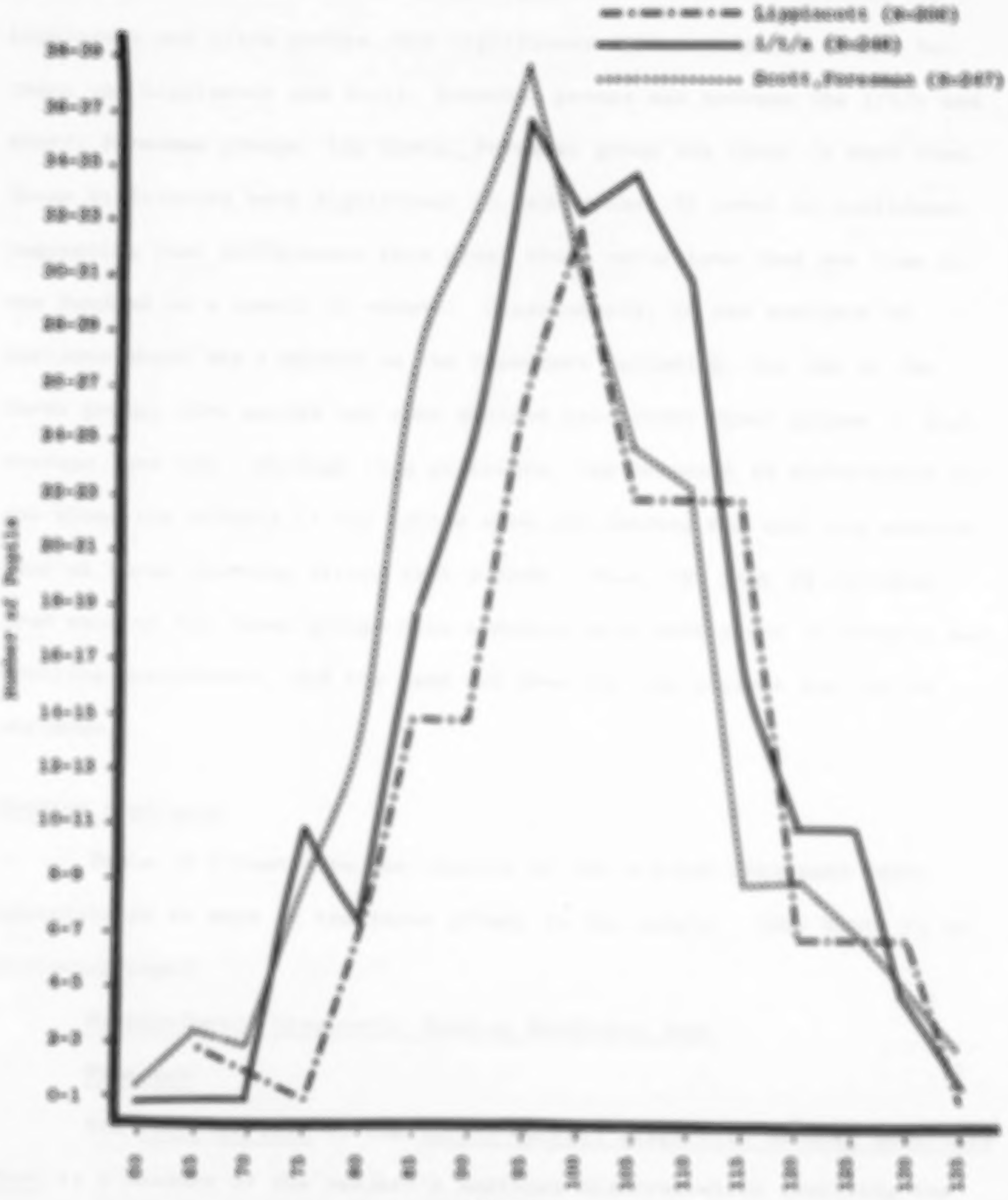

Figure 1 Frequency distributions of IQ scores on the Pintner-Cunningham Primary Test of the three treatment groups.

analyses revealed that there was no significant difference between the Lippincott and i/t/a groups, but significant differences did occur, between the Lippincott and Scott, Foresman groups and between the i/t/a and Scott, Foresman groups; the Scott, Foresman group was lower in each case. These differences were significant at beyond the .01 level of confidence, suggesting that differences this great could occur less than one time in one hundred as a result of chance. Consequently, in the analysis of variance which was computed on the dependent variables, the IQs of the three groups were pooled and then divided into three equal groups -- high, average, and low. Through this procedure, the observed IQ differences did not cloud the effects of the system upon the reading and spelling achievement of those learning within that system. Thus, the high IQ children from each of the three groups were compared with each other in reading and spelling achievement, and the same was done for the average and low IQ children.

Reading readiness

Table 14 illustrates the results of the initial readiness tests administered to each of the three groups in the sample. (See Table 14 on following page).

Murphy-Durrell Diagnostic Reading Readiness Test

Phonemes

The Phonemes Test of the Murphy-Durrell Diagnostic Reading Readiness Test is a measure of the subject's auditory discrimination involving the identification of initial and final sounds of words. As can be seen in Table 14, the mean raw scores were 24.6 for the Lippincott group, 21.8 for the i/t/a group, and 24.1 for the Scott, Foresman group. The obtained

TABLE 14

MEANS AND STANDARD DEVIATIONS OF READING READINESS SCORES OF THE THREE TREATMENT GROUPS

	Lippincott (Carle Place) N=208		L/T/a (East Meadow) N=248		Scott, Foresman (Uniondale) N=207	
Test[a]	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	34.6	12.6	31.8	12.2	34.1	12.7
2	22.0	5.6	19.2	6.6	20.2	6.2
3	18.1	5.7	16.6	6.6	16.2	6.2
4	41.0	11.1	35.6	12.6	36.5	12.1
5	12.3	4.5	10.2	3.9	10.8	3.7
6	13.3	4.6	12.4	7.2	12.1	6.9
7	16.5	5.6	13.7	6.3	13.1	5.3
8	8.9	2.6	8.6	2.5	8.8	2.5
9	8.5	2.2	8.2	2.6	8.2	2.2
10	67.7	12.1	62.4	13.7	62.8	12.0

KEY[a]

Test	
1	Murphy-Durrell Phonemes
2	Murphy-Durrell Letter Names - Capitals
3	Murphy-Durrell Letter Names - Lower-case
4	Murphy-Durrell Letter Names - Capital & Lower-case letters combined
5	Murphy-Durrell Learning Rate
6	Thurstone Pattern Copying
7	Thurstone Identical Forms
8	Metropolitan Word Meaning
9	Metropolitan Listening
10	Metropolitan Total Readiness Score

standard deviations for each of these three groups was approximately equal. Critical ratios were computed between each of the possible pairs of mean scores on this measure of readiness. The results indicate that the Lippincott group scored significantly higher than the i/t/a group in Phonemes, with significance at the .05 level and approaching the .01 level of confidence. Between the Lippincott and Scott, Foresman groups, the Lippincott group scored significantly higher, well beyond the .01 level of confidence. The i/t/a group also scored significantly higher than the Scott, Foresman group at well beyond the .01 level of confidence. The differences between the Lippincott and Uniondale groups and the i/t/a and Uniondale groups would be expected on the basis of the significant differences in IQ among these three groups, with both the Lippincott and i/t/a groups significantly higher than the Scott, Foresman group. However, the observed difference between the Lippincott and i/t/a groups would not be expected on the basis of intelligence alone, and would suggest the possibility that these differences resulted from differences in the kindergarten readiness programs which the children had experienced prior to first-grade entrance.

Letter Names

The Letter Names Test measures the pupil's ability to recognize both capital and lower-case letter names. Raw score means were reported separately for capital letters, lower-case letters, and the combined total score.

On Capital Letter Names, the respective means were 22.0, 19.2, and 20.2 for the Lippincott, i/t/a, and Scott, Foresman groups. The Lippincott, group scored significantly higher than the i/t/a or Scott, Foresman groups at beyond the .01 level of confidence. The difference observed between the i/t/a and Scott, Foresman groups was not significant at the .05 level, although the mean of the Scott, Foresman group was slightly higher. This, again, suggested differences among the readiness programs of the Lippincott group as compared with the Scott, Foresman and i/t/a groups.

On Lower-case Letter Names, the means respectively for the Lippincott, i/t/a, and Scott, Foresman groups were 19.1, 14.6, and 16.2. The average raw score of the Lippincott group was significantly higher than the means of both the i/t/a and Scott, Foresman groups, with the observed differences significant at a level well beyond the .01 level of confidence. The Scott, Foresman group, however, was significantly higher than the i/t/a group in recognizing Lower-case Letter Names at the .01 level of confidence.

The combined total score for capital and Lower-case Letter Names reveals, again, the superiority of the Lippincott group, as compared to the i/t/a and Scott, Foresman groups. The Lippincott group was significantly higher than the Uniondale or Scott, Foresman groups, with the results significant beyond the .01 level of confidence. The Scott, Foresman group's mean was significantly higher than that of the i/t/a group at the .05 level of confidence.

Because of these differences, an examination was made of the kindergarten readiness programs in each of the three school systems from which the samples were drawn. From this examination, it was determined, in fact, that the kindergarten readiness program in Carle Place, where the Lippincott system was used, included weekly instruction in capital and lower-case letter names on a regular basis. In Uniondale and East Meadow, where the Scott, Foresman and i/t/a systems were used respectively, their kindergarten programs allowed instruction in these areas at the teachers' options; and this instruction, when it occurred, was on an incidental rather than a planned basis.

Learning rate

The Learning Rate Test measures the subject's ability to learn new words. The results, as shown in Table 14, indicate respective means of 12.3, 10.2, and 10.9 for the Lippincott, i/t/a, and Scott, Foresman groups. Again, the average score of the Lippincott group was significantly higher than the other two groups at well beyond the .01 level of confidence. The slight, observed difference between the i/t/a and Scott, Foresman groups, with the Scott, Foresman group slightly higher, was not significant at the .05 level.

Thurstone Tests

Visual discrimination ability was measured by the Thurstone Pattern Copying and Identical Forms tests. In the Pattern Copying Test, the subject has to draw additional lines to complete patterns. In the Identical Forms Test, the subject marks one form out of five choices which is identical to the first form in the row.

In Pattern Copying, the means of the three groups were relatively similar. A significant difference between means was observed only between the Lippincott and Scott, Foresman groups, and this difference was barely significant at the .05 level of confidence. No significant difference was observed between the means of the Lippincott and i/t/a groups and between the i/t/a and Scott, Foresman groups.

The Identical Forms Test, however, reveals extreme differences between means. The Lippincott group was significantly higher on this subtest than either the i/t/a or Scott, Foresman groups. In addition, the i/t/a group had a significantly higher mean than the Scott, Foresman group. All of these differences were significant at beyond the .01 level of confidence.

Metropolitan Readiness Tests

All tests of the revised edition of the Metropolitan Readiness Tests were administered, as requested by the publisher of this test; but only the first two subtests, Word Meaning and Listening, and the total score were utilized in this study. Both the Word Meaning and Listening subtests were used to determine the pupil's auditory language ability.

Word Meaning

On Word Meaning, respective means for the Lippincott, i/t/a, and Scott, Foresman groups were 9.9, 9.6, and 8.8. Standard deviations of the three groups were similar. An analysis of these results indicated no significant difference between the Lippincott and i/t/a groups. Significant differences were observed between the i/t/a and Scott, Foresman groups and between the Lippincott and Scott, Foresman groups. In both cases, the Scott, Foresman group scored significantly lower, with the difference between means significant at the .01 level of confidence.

Listening

On the Listening subtest, the resulting means were relatively similar, ranging between a low of 9.2 for the i/t/a group, to a high of 9.5 for the other two groups. Comparisons between each of the pairs of means revealed that no significant differences resulted at the .05 level.

Total readiness score

On the total score of the Metropolitan Readiness Tests, which was the sum of raw scores for the six subtests, the Lippincott, i/t/a, and Scott, Foresman groups had mean scores of 67.7, 62.4, and 62.8 respectively. Again, as on many of the other readiness measures, the Lippincott group scored significantly higher than the other two. These differences were, also, significant at the .01 level of confidence or beyond. No significant difference was observed between the i/t/a and Scott, Foresman groups. There was a difference of only .4 between the means of these two groups, in favor of the Scott, Foresman group.

Summary Statement

In almost all aspects of the readiness tests, administered at the beginning of the school year, prior to any instruction, the group taught to read in the Lippincott series was significantly superior to the two other groups. The only exceptions were the Thurstone Pattern Copying and the Word Meaning and Listening subtests of the Metropolitan Readiness Tests. In all readiness tests, but one, the Lippincott group was significantly superior to the Scott, Foresman group; the only exception occurring on the Listening subtest of the Metropolitan Readiness Tests. Between the i/t/a and Scott, Foresman groups, significantly superior readiness scores in favor of the i/t/a group occurred on the following tests: Murphy-Durrell Phonemes, Thurstone Identical Forms, and Metropolitan Word Meaning. Significant differences in favor of the Scott, Foresman group were observed in Lower-case Letter Names and capital and lower-case Letter Names combined of the Murphy-Durrell Diagnostic Readiness Test. On the other readiness tests, no significant differences were observed between the i/t/a and Scott, Foresman group. These results indicated that the i/t/a and Scott, Foresman groups were relatively comparable in reading readiness at the time the study was initiated. However, it was also relatively evident the group instructed in the Lippincott system was significantly superior to the other two groups in their readiness for reading, as measured on the initial tests. This can be accounted for by two factors: The first is that the Lippincott group had the highest mean IQ of any of the three groups of the study, and the kindergarten readiness program in the Carle Place district, from which the Lippincott sample was drawn, had the most structured readiness program in terms of providing specific, regularly scheduled instruction in some of the variables measured in the readiness tests. This difference was considered in evaluating and interpreting the results obtained in reading and spelling that were the outcome variables of this study.

Results of the Teachers' Logs

Instructional time

At the commencement of the study, all teachers were requested to spend approximately two and one-half hours per day in direct reading instruction and supportive activities. Direct reading instruction was defined as time spent by the teacher in direct supervision of learning activities in reading. Supportive activities were defined as time spent in the teaching of reading other than the basal reader. The teachers kept time logs of activities, skills, and materials used in instruction. Validation of the logs was determined by observation of the research staff who compared their logs, kept when visiting the classroom, with the teachers' logs. From an analysis of the teacher's log, the time spent in direct reading instruction plus the time given to supportive activities were combined to determine whether the total time allotment was adhered to approximately. Since time schedules tend to vary from day to day in a teaching situation, the results of the time spent in direct and supportive reading instruction were averaged for a total week. A week was chosen at random in the fall and in the spring for a comparison of the three treatment groups in the area of time spent in reading instruction. The table for the fall follows:

TABLE 15

SIGNIFICANCE OF THE DIFFERENCES BETWEEN TREATMENTS IN AVERAGE NUMBER OF MINUTES SPENT PER DAY BY TEACHERS IN DIRECT READING INSTRUCTION AND SUPPORTIVE ACTIVITIES

FALL, 1964

	N	Mean	S.D.	t
Lippincott	8	150	35	.07
i/t/a	9	151	9	
Lippincott	8	150	35	.37
Scott, Foresman	9	155	7	
i/t/a	9	151	9	1.00
Scott, Foresman	9	155	7	

Table 15 reveals that the teachers in each of the three treatment groups adhered fairly closely to the time allotment. In all of the groups, the time spent in direct and supportive instruction in reading was approximately 150 minutes. Slight differences existing among the three groups were not statistically significant at the .05 level. A week was, also, chosen at random in the spring and the time spent in reading activities was averaged for each teacher in each of the three treatment groups for an entire week. The table for spring follows:

TABLE 16

SIGNIFICANCE OF THE DIFFERENCES BETWEEN TREATMENTS IN AVERAGE NUMBER OF MINUTES SPENT PER DAY BY TEACHERS IN DIRECT READING INSTRUCTION AND SUPPORTIVE ACTIVITIES

SPRING, 1965

	N	Mean	S.D.	t
Lippincott	8	123	16	3.65**
i/t/a	9	156	19	
Lippincott	8	123	16	3.72**
Scott, Foresman	9	152	14	
i/t/a	9	156	19	.48
Scott, Foresman	9	152	14	

**Significant at .01 level

As can be observed in Table 16, the Early-to-Read i/t/a and the Scott, Foresman groups spent approximately the same time on reading activities for that week. These times, also, approximate the 150 minutes expected time allotment. In the Lippincott group, however, for that week, teachers spent slightly over two hours in reading activities, which was approximately one-half hour less than that expected. The amount of time spent by the Early-to-Read i/t/a group teachers and the Scott, Foresman group teachers was significantly higher than that spent by the Lippincott teachers at the .01 level of confidence. No significant difference existed between the Early-to-Read i/t/a and Scott, Foresman groups.

In the fall, there was very little variability in the amount of time spent per day in reading instruction. In the spring, however, the Lippincott group tended to vary much more widely than the Scott, Foresman or Early-to-Read i/t/a groups. This reflects the fact that on some days the Lippincott teachers spent very little time on reading in that given week, and on other days, instruction in reading took place for more than the two and one-half hours expected. Nevertheless, in total, the time spent by the Lippincott teachers was equivalent to that of the two other treatment groups. The results of the logs for the two weeks selected were validated against the observer's reports. In each case, the logs kept by the teacher were approximately equivalent to the reports of the research assistants. Research assistants, however, reported slightly more time in reading instruction spent in the Lippincott group than the teachers themselves reported. Hence, it is possible that the Lippincott teachers were slightly less accurate in keeping their logs during that week. In general, time spent in reading instruction would not seem to be a major factor effecting the final results.

Classroom organization

In the training sessions which preceded the program of instruction in each of the treatment groups, emphasis was placed on adapting instruction to the learning needs of children, with specific attention directed toward adapting the level of materials used to the abilities of the children and recognition of differences in rates of progress.

This would be reflected in the amount of grouping used in the classroom. From the teacher's log, it was possible to determine the number of groups with which a teacher worked within the classroom. Table 17 illustrates the classroom organization of the teachers in the fall and spring.

TABLE 17

NUMBER OF READING GROUPS BY PERCENTAGE OF TEACHERS WITHIN EACH TREATMENT

FALL AND SPRING 1964-65

	Number of Reading Groups							
	1		2		3		4	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
Lippincott N=8 teachers	0	0	0	0	100	63	0	37
i/t/a N=9 teachers	0	0	0	0	100	33	0	67
Scott, Foresman N=9 teachers	0	0	0	0	100	88	0	12

In Table 17 it can be seen that in the fall, in each of the three treatment groups, all of the teachers organized their classes into three groups. However, in the spring, teachers tended toward more individualization of instruction, as some of the teachers organized their classes using more than three groups. For the Lippincott group, 63% of the teachers were still using a three-group system of organization in the spring, while 37%, or approximately more than one-third of the teachers utilized four or more groups within their classroom structure. In the Scott, Foresman group, 88% of the teachers continued their fall utilization of a three-group classroom organization, while 12% of the teachers divided their classes further into four groups or more. For the *Early-to-Read i/t/a* group, one-third of the teachers continued with the three-group system that they had utilized in the fall, whereas two-thirds of the teachers individualized their classes to the point where four groups or more were utilized. Examination of the above results suggests that the Scott, Foresman teachers were less flexible in their attempts to meet individual needs with a small percentage of teachers showing greater individualization later in the year than they had exhibited earlier in the year. The Lippincott teachers were slightly more flexible than those teaching by the Scott, Foresman system, yet a little less than two-thirds of the teachers continued with a classroom organization similar to that which they had used in the fall, whereas slightly more than one-third of the teachers further individualized their classes. The *Early-to-Read i/t/a* group exhibited the greatest amount of flexibility in classroom organization, so although 100% of these teachers were using three groups in the fall, only one-third of these teachers were still utilizing three groups by the spring semester, while two-thirds of the teachers felt that the different rates of progress exhibited by their children required more than three groups for effective handling.

Instructional levels in reading reported by teachers

Teachers were requested, in May, to report on the names and levels of the books in which the children were being instructed as of May 14. Hence instructional level, as defined to the teachers was that reader level at which a pupil is receiving instruction under the teacher's direction. Table 18 illustrates the percentage of pupils receiving instruction at each reader level of the Scott, Foresman group.

TABLE 18

PERCENTAGE OF PUPILS AT EACH INSTRUCTIONAL READER LEVEL OF THE SCOTT, FORESMAN POPULATION

Reader Level	Percentage of Pupils N=239
2^1	3.8
1^2	54.
1	9.
Primer	18.2
Jr. Primer	12.4
PP3	.6
	100.0%

As can be seen in Table 18, 54%, or slightly more than half of the children were reading at the 1^2 level, which would normally be expected by the end of first-grade. Approximately 42% of the children were reading below expected first-grade standards for the middle of May, as they were reading in the junior primer or lower. Approximately 30% of the children were reading at a primer level or below, while less than 1% of the children were reading below a junior primer level. Approximately 4% of the children in the Scott, Foresman group were being instructed in books rated as being more difficult than first-grade level.

Table 19 indicates the percentage of pupils at each reader level of the Lippincott population.

TABLE 19

PERCENTAGE OF PUPILS AT EACH INSTRUCTIONAL READER LEVEL OF THE LIPPINCOTT POPULATION

Reader Level	Percentage of Pupils N=203
2^2	4
2^1	13
1^2	30
1^1	36
Primer	10
Preprimer	7
	100%

Table 19 shows that 36% of the children were reading in the 1^1 book, 83% were reading at a 1^2 level or lower, 10% of the Lippincott children were reading at a primer level, and 7% were reading in preprimers. Thirty per cent of this group were reading at the 1^2 level, and approximately 17% of this group were reading at levels beyond high first-grade. Five of the total of eight teachers introduced the Science Research Associates (SRA) Reading Laboratory at the beginning of May for those children who had completed the 1^2 book of the Lippincott series. This material was used as individual, independent, and/or recreational reading with the top and middle reading groups. The number of children using SRA in each classroom ranged from 14 to 21. The reading levels of these children ranged from 1^2 to

3^1, according to the SRA measurements.

Comparing the results of the Lippincott and Scott, Foresman groups in terms of teacher reported instructional levels suggests that the Lippincott group had a higher number of children reading beyond a 1^2 level, but the Lippincott group, also, had a higher percentage of children reading at levels below primer. This suggests that the Lippincott system may have some advantages for some children and may be disadvantageous for other.

Table 20 presents the percentage of children receiving reading instruction at each of the reader levels of the i/t/a sample.

TABLE 20

PERCENTAGE OF PUPILS AT EACH INSTRUCTIONAL READER LEVEL OF THE i/t/a POPULATION

Reader Level	Percentage of Pupils N=239
3^1 Book 7 (Traditional orthography)	48
2^2 (Traditional orthography readers)	10
2^1 (Traditional orthography readers)	14
Book 6 (3.4 i/t/a mean readability)	14
Book 5 (2.9 i/t/a mean readability)	7.8
Book 4 (2.4 i/t/a mean readability)	5
Book 3 (2.0 i/t/a mean readability)	.8
Book 2 (1.9 i/t/a mean readability)	.4
	100.0%

From Table 20, it can be seen that 24% of the i/t/a children had completed the last book of the series, where transition is made from reading in i/t/a to reading in traditional orthography. For this 24%, books in traditional orthography used for instruction after transition ranged from a 2^1 to a 2^2 level. Forty-eight per cent of the children in the sample were reading in Book 7. The readability level of Book 7, which is written almost completely in traditional orthography, is 3.0 by the Spache (1953) formula. Over 21% of the Early-to-Read i/t/a sample were reading in Book 5 or 6 of the Early-to-Read i/t/a series, which ranges in readability from a 2.9 to a 3.4 level in i/t/a. Five per cent of the children were reading in Book 4 which has a mean readability level of 2.4. A little over 1% of the sample were reading in books below the Book 4 level, where the mean readability levels would be 2.0 or lower. The readability levels, as determined by formula for Book 6 or readers lower in the Early-to-Read i/t/a series need to be interpreted cautiously, since the Spache formula was prepared for primary books written in traditional orthography. Book 7, however, written almost completely in traditional orthography has a 3.0 mean readability level which is likely to be more valid.

Comparing the results of the Lippincott, Scott, Foresman, and i/t/a groups with respect to the instructional levels reported by teachers suggests the following differences for the highest and lowest achieving pupils: Seventy-two per cent of the Early-to-Read i/t/a children were reading at a 2^1 level or higher, while only 4% of the Scott, Foresman children were reading beyond the 1^2 level, and 17% of the Lippincott children exceeded a level of 1^2. At the bottom end, 7% of the Lippincott group were reading at pre-primer level, less than 1% of the Scott, Foresman group were reading at a pre-primer level or lower, and less than 1% of the Early-to-Read i/t/a group were below primer level. These results would suggest that the Early-to-Read i/t/a group had the greatest number of children whose reading achievement exceeded normal grade expectancy, while Scott, Foresman had the lowest number. Among the least successful readers, the Scott, Foresman and Early-to-Read i/t/a groups had very insignificant percentages (less than one), while the Lippincott group had the highest percentage of the low achievers in reading.

Evaluation of Teacher Effectiveness

The 27 project directors participating in the Cooperative Research Program's first-grade reading study agreed to rate teachers on a descriptive scale and criteria provided by the Coordinating Center at the University of Minnesota. Teachers were rated in three categories on a five-point scale. These categories were: (a) class structure, (b) extent of class participation, and (c) awareness of and attention paid to individual needs of pupils. In addition, an overall rating determining the teachers' competence was derived from the foregoing three categories. A member of the research staff was assigned to each of the districts participating in the study to rate the teachers on the categories given and according to the criteria described in each of these categories. Prior to the rating of the teachers, a meeting of the raters was held at which common grounds for evaluating the criteria within categories was covered. An attempt was made to ensure that each of the observers would rate the teachers in terms of the criteria given and not allow their own personal feelings to effect their evaluations.

Table 21 illustrates the percentage of teachers within each of the treatment groups for each rating criterion within each category. The rating scale and criteria are described in Appendix B.

TABLE 21

PERCENTAGE OF TEACHERS IN EACH TREATMENT GROUP AS RATED ON A FIVE-POINT SCALE IN CLASS STRUCTURE, CLASS PARTICIPATION, ATTENTION TO INDIVIDUAL NEEDS, AND OVER-ALL TEACHER COMPETENCE

	Class Structure					Class Participation					Attention to Individual Needs					Over-all Teacher Competence				
Rating	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Treatment																				
Lippincott	0	0	24	38	38	0	0	25	25	50	0	0	12	38	50	0	0	25	25	50
i/t/a	0	0	0	22	78	0	0	11	67	22	0	0	0	67	33	0	0	0	67	33
Scott, Foresman	0	0	44	34	22	0	0	22	67	11	0	0	34	33	33	0	0	34	44	22

As can be seen from Table 21, none of the observers rated any of the teachers in any of the three districts as being incompetent. Under the category of class structure, in the district utilizing the Scott, Foresman basal system, 56% of the teachers fell in the highest two ratings, indicating a great deal of structure with clear direction and expectations. Slightly over 44% of the Scott, Foresman teachers were classified as average in structuring their classes, with some degree of vagueness present. This is contrasted to the East Meadow district, where i/t/a was utilized. All of the teachers fell into the top two classifications. Under class structure, 78% of them obtaining the highest rating of four. This indicates that, in general, the i/t/a teachers are well organized and clear in assigning tasks; directions and expectations, for almost all the teachers are clearly spelled out in detail. A little over three-fourths of the Lippincott teachers obtained the highest two ratings in structuring thier classes. Twenty-four per cent of the teachers were considered average in structuring their classes. In summary, then, the i/t/a teachers rated higher in structuring their classrooms than did the Lippincott or Scott, Foresman teachers. The ratings in this area for the Lippincott teachers were also somewhat higher than those of the Scott, Foresman teachers.

The second category evaluated was the extent of class participation. A rating of two would be considered average in this respect. Ratings of three or four would suggest moderate to high participation on the part of most of the children most of the time, whereas ratings of zero or one would indicate that the class is, generally, unresponsive, with high variability in participation, and very few children actually participating in classroom activities. In this category, the i/t/a teachers, again, obtained the highest ratings, with 89% of their teachers considered to be above average in developing good classroom participation. Although both Scott, Foresman and Lippincott teachers have lower ratings than the i/t/a teachers attained, approximately 75% of both the Scott, Foresman and Lippincott teachers also were rated as being above average in developing class participation. In summary, then, teachers were rated by the research staff as developing good classroom participation, with teachers drawing most of the children into classroom activities.

The degree to which the teachers seemed to be aware of and attended to the individual needs of pupils and individualized instruction was, also, evaluated and rated. Again, in this category, two was considered average, zero and one definitely below average, and three and four definitely above average in degree of individualization. An examination of the ratings attained indicates that i/t/a teachers, again, obtained the highest ratings, all of the teachers rating above average in this category. Lippincott teachers, also, were given exceedingly high ratings, with 88% of these teachers rating above average. Of the Scott, Foresman teachers, 66% were rated as being above average, while 34% were considered to be average in individualizing their programs according to pupil needs.

Ratings obtained in the three categories were averaged for the teachers within each treatment group, to arrive at an over-all rating of teacher competence. A rating of one would be considered poor, two would be considered adequate, three would be considered good, and four would be considered excellent. All of the i/t/a teachers were evaluated as good or excellent by the observer, with 67% falling into the rating of good, and 33% falling into the rating of excellent. In the Lippincott group, 50% of the teachers were considered excellent, 25% as good, and 25% as adequate. The ratings obtained for the Scott, Foresman teachers indicate that 22% were rated as excellent, 44% as good, and 33% as being adequate.

In summary, then, the i/t/a teachers obtained the highest ratings. The Lippincott teachers attained high ratings, not quite as high as the i/t/a teachers although they had the highest percentage of excellent ratings. The Scott, Foresman teachers rated lowest, although two-thirds of these teachers rated in the good or excellent category. In terms of the description of the teacher sample, described previously in another section of this report, these results are rather interesting since the i/t/a teachers tended to be the youngest and had the least amount of teaching experience prior to the commencement of the study. To some extent, this may help compensate for some of the factors where differences were observed above in the areas of teachers' ages, pre-first-grade school experience, and total previous teaching experience.

Results of Post-instruction Tests on Subsamples

The following post-instruction word recognition and oral reading measures were individually administered, in May, to a very small subsample of randomly selected pupils within each treatment group: Gilmore Oral Reading Test, Gates Word Pronunciation Test, Karlsen Phonemic Word Test, and Phonetically Regular Words Oral Reading Test (see Appendix C for copies of word recognition measures). All tests were administered in traditional orthography. Critical ratios were computed between each of the possible pairs of mean scores on each of the outcome variable measures to determine whether the observed differences between means were significant or could have occurred by chance.

Gilmore Oral Reading Test

The Gilmore Oral Reading Test consists of a series of paragraphs of increasing difficulty which the pupil is asked to read aloud at sight. Two comprehension questions were asked after each paragraph but were disregarded in the scoring of the test, in accordance with the procedures adopted in the Cooperative Research Program's first-grade reading study. Two parts of the Gilmore test were reported in this study: Accuracy, which was recorded as a grade equivalence score, and Rate, recorded as the number of words read per minute. The score for Accuracy was ten points for each paragraph read correctly, less the number of errors. Hesitations and repetitions were not counted as errors.

As can be seen from Table 22, there was no significant difference between the Scott, Foresman and Lippincott groups and between the Lippincott and i/t/a groups. However, the results clearly indicate that the i/t/a group scored significantly higher than the Scott, Foresman group in Accuracy. The difference between mean grade equivalency scores was significant at the .01 level of confidence.

TABLE 22

SIGNIFICANCE OF THE DIFFERENCES BETWEEN
MEAN GRADE EQUIVALENT SCORES IN ACCURACY
ON THE GILMORE ORAL READING TEST OF
RANDOMLY SELECTED PUPILS WITHIN THE THREE TREATMENTS

Treatment	N	Mean	S.D.	t	Sig.
Lippincott	16	3.1	1.4	1.15	-
Early-to-Read i/t/a	18	3.6	1.5		
Lippincott	16	3.1	1.4	1.67	-
Scott, Foresman	17	2.3	1.1		
Early-to-Read i/t/a	18	3.6	1.5	2.91	.01
Scott, Foresman	17	2.3	1.1		

Comparisons among the three groups on the average number of words read per minute (Rate Score) on the Gilmore test are presented in Table 23.

TABLE 23

SIGNIFICANCE OF THE DIFFERENCES BETWEEN
MEAN NUMBER OF WORDS READ PER MINUTE (RATE)
ON THE GILMORE ORAL READING TEST OF
RANDOMLY SELECTED PUPILS WITHIN THE THREE TREATMENTS

Treatment	N	Mean	S.D.	t	Sig.
Lippincott	16	69.3	27.4	1.97	-
Early-to-Read i/t/a	18	52.1	22.9		
Lippincott	16	69.3	27.4	2.93	.01
Scott, Foresman	17	45.9	17.0		
Early-to-Read i/t/a	18	52.1	22.9	0.92	-
Scott, Foresman	17	45.9	17.0		

The average number of words read per minute was 69.3, 52.1, and 45.9, respectively for the Lippincott, i/t/a, and Scott, Foresman groups. The Lippincott group read faster than the i/t/a group, but the difference between the groups was not significant. The Lippincott group, however, read significantly faster than the Scott, Foresman children, with the difference in rate significant beyond the .01 level of confidence. There was no difference between the i/t/a and Scott, Foresman groups.

Gates Word Pronunciation Test

The Gates Word Pronunciation Test consists of a list of forty high frequency words, graded in order of difficulty, which the pupil is asked to read aloud. The words are not controlled with respect to spelling regularity. The score is the number of words read correctly.

From Table 24 it can be seen that the mean score obtained by the Scott, Foresman group was significantly lower than the mean scores of both the Lippincott and i/t/a groups, with the observed differences significant beyond the .01 level of confidence. No significant difference was observed between the Lippincott and i/t/a subsamples. The results suggest that the Lippincott and i/t/a pupils were superior to the Scott, Foresman pupils with respect to recognizing a larger number of words which occur frequently in primary grade reading materials.

TABLE 24

SIGNIFICANCE OF THE DIFFERENCES BETWEEN
MEAN SCORES ON THE GATES WORD PRONUNCIATION TEST OF
RANDOMLY SELECTED PUPILS WITHIN THE THREE TREATMENTS

Treatment	N	Mean	S.D.	t	Sig.
Lippincott	16	20.4	9.8	1.13	-
Early-to-Read i/t/a	18	23.9	7.7		
Lippincott	16	20.4	9.8	3.81	.01
Scott, Foresman	17	9.9	5.1		
Early-to-Read i/t/a	18	23.9	7.7	6.35	.01
Scott, Foresman	17	9.9	5.1		

Karlsen Phonemic Word Test

The Karlsen Phonemic Word Test is unpublished and has no norms. The test consists of a list of forty words which have a high degree of spelling regularity. The pupil is asked to read the words aloud until he has missed five consecutive words. The score is the number of words read correctly.

As can be seen in Table 25, the Lippincott and i/t/a children attained significantly higher mean scores than the Scott, Foresman group. The difference was significant at the .01 level of confidence in each case. No significant difference was observed between the Lippincott and i/t/a groups. The results, again, demonstrate that the i/t/a and Lippincott children were superior to the Scott, Foresman children in recognizing or analyzing regularly spelled words.

TABLE 25

SIGNIFICANCE OF THE DIFFERENCES BETWEEN
MEAN SCORES ON THE KARLSEN PHONEMIC WORD TEST OF
RANDOMLY SELECTED PUPILS WITHIN THE THREE TREATMENTS

Treatment	N	Mean	S.D.	t	Sig.
Lippincott	16	24.1	11.0	.64	-
Early-to-Read i/t/a	18	26.3	8.0		
Lippincott	16	24.1	11.0	5.10	.01
Scott, Foresman	17	7.8	6.9		
Early-to-Read i/t/a	18	26.3	8.0	7.38	.01
Scott, Foresman	17	7.8	6.9		

Phonetically Regular Words Oral Reading Test

This test consists of a list of forty phonetically regular words graded according to approximate level of difficulty by vowel sounds used. The pupil is asked to read the words aloud. The score is the number of words read correctly.

As can be seen from Table 26, the respective means for the Lippincott, i/t/a, and Scott, Foresman children were 20.8, 24.8, and 3.8.

TABLE 26

SIGNIFICANCE OF THE DIFFERENCES BETWEEN MEAN SCORES
ON THE PHONETICALLY REGULAR WORDS ORAL READING TEST
FOR RANDOMLY SELECTED PUPILS WITHIN THE THREE TREATMENTS

Treatment	N	Mean	S.D.	t	Sig.
Lippincott	16	20.0	5.7	1.08	-
Early-to-Read i/t/a	18	24.8	17.9		
Lippincott	16	20.0	5.7	8.94	.01
Scott, Foresman	17	3.8	4.5		
Early-to-Read i/t/a	18	24.8	17.9	4.81	.01
Scott, Foresman	17	3.8	4.5		

The Lippincott and i/t/a children were, again, very significantly superior to the Scott, Foresman children, with significance beyond the .01 level of confidence. The i/t/a and Lippincott pupils did not differ significantly from each other. The superiority of both the Lippincott and i/t/a children over the Scott, Foresman children, again, suggests differences in the reading programs of the three populations. The Scott, Foresman basal reader system does not introduce vowels in its first-grade program, whereas the Lippincott and i/t/a readers introduce all the vowel sounds in the first-year program. Hence, the Scott, Foresman children would be at a disadvantage in analyzing unfamiliar words containing vowels.

To determine whether the observed differences in the results of the oral reading and word recognition tests could have been due to differences in the mean IQ scores of the pupils within each subsample, an examination was made of the Pintner-Cunningham test results of the three samples.

TABLE 27

SIGNIFICANCE OF THE DIFFERENCES BETWEEN MEAN IQ SCORES
ON THE PINTNER-CUNNINGHAM PRIMARY TEST OF
RANDOMLY SELECTED PUPILS WITHIN THE THREE TREATMENTS

Treatment	N	Mean	S.D.	t	Sig.
Lippincott	16	105.8	15.3	0.51	-
Early-to-Read i/t/a	18	103.2	14.6		
Lippincott	16	105.8	15.3	2.04	.05
Scott, Foresman	17	95.4	14.0		
Early-to-Read i/t/a	18	103.2	14.6	1.64	-
Scott, Foresman	17	95.4	14.0		

As can be seen from Table 27, there was no significant difference between the i/t/a and Lippincott pupils in mean IQ score. However, the Lippincott group had a significantly higher mean IQ score than the Scott, Foresman pupils, with marginal significance observed at the .05 level of confidence. Between the Lippincott and i/t/a pupils, no significant difference between mean IQ scores was observed. The fact that the Lippincott children had a significantly higher mean IQ score than the Scott, Foresman pupils is a factor that should be considered in interpreting the results on the oral reading and word recognition tests for these two groups.

Summary statement

The pattern of results is very consistent on the oral reading and word recognition tests administered at the end of the experimental period to a randomly selected subsample within each of the treatment groups. The results indicate that the Lippincott and i/t/a pupils were comparable, and that both were significantly superior to the Scott, Foresman pupils on every measure except two. On Gilmore Accuracy no significant difference was observed between the Lippincott and Scott, Foresman pupils, but a significant difference was found between the i/t/a and Scott, Foresman groups in Accuracy, in favor of the i/t/a group. On Gilmore Rate, no significant difference was observed between the i/t/a and Scott, Foresman pupils. However, the Lippincott group was significantly higher than the Scott, Foresman group on Rate. Between the Lippincott and i/t/a pupils there was no significant difference on any of the measures, although the i/t/a group's mean scores were higher on every test but one, which was Gilmore Rate. Generally, the results suggest that the superiority of the i/t/a and Lippincott subsamples over the Scott, Foresman group may be attributed to differences in the instructional basal programs in which the children in the three populations were taught to read. Children in the i/t/a and Lippincott groups were instructed by procedures which are considerably more analytic than the procedures by which the Scott, Foresman children were instructed. The instructional procedures of the i/t/a and Lippincott basal systems place heavy emphasis upon word structure and the phonic elements of words. Moreover, the vocabulary load is substantially higher in the i/t/a and Lippincott series than in the Scott, Foresman series. Thus, it would be expected that both the i/t/a and Lippincott pupils would possess skills which enable them to recognize and analyze a considerably greater number of words than the Scott, Foresman pupils.

Results and Discussion: Reading Achievement

The primary purpose of this study was to determine whether any differential effects occurredin reading achievement as a result of the use of three basal reader systems containing discreet features. Since intelligence would be considered an important variable, comparisons among the three reader systems were accomplished for high, middle and low IQ children. This was done by obtaining the total distribution of IQs for the total number of subjects in each of the three treatment groups, and dividing this distribution into thirds (high, middle, and low). The comparisons among the three treatments were then made for children of high, average, and low intelligence. Since the children did not qualify as independent observations, and since children in the same class received the same type of instruction, the class was used as the unit of observation. In addition to the comparison of reading achievement made by intelligence, the samples for each of the three treatment groups in each of the three categories of intelligence were divided by sex. The comparisons among the three systems were then made for boys of high, average, and low intelligence and for girls of high, average, and low intelligence. The high, average, and low IQs were sorted by sex and treatment. Table 28 reveals the number of subjects and mean IQ scores in each IQ category by sex and treatment.

TABLE 28

NUMBER AND MEAN IQ SCORES OF SUBJECTS CLASSIFIED BY IQ LEVEL, SEX, AND TREATMENT

IQ Level	Group	N	Mean
High IQ	Lippincott boys	36	115.6
	Lippincott girls	38	117.5
	i/t/a boys	32	116.9
	i/t/a girls	55	115.6
	Scott, Foresman boys	22	113.6
	Scott, Foresman girls	35	117.1
Average IQ	Lippincott boys	37	100.8
	Lippincott girls	30	101.6
	i/t/a boys	48	100.6
	i/t/a girls	33	101.5
	Scott, Foresman boys	32	100.0
	Scott, Foresman girls	39	101.7
Low IQ	Lippincott boys	34	86.6
	Lippincott girls	14	85.8
	i/t/a boys	43	86.7
	i/t/a girls	27	87.7
	Scott, Foresman boys	64	86.7
	Scott, Foresman girls	37	87.4

Composite reading score

The first comparison among the children learning in each of the three basal reader systems was on a composite reading score, since this would be considered as a single measure of over-all reading achievement, and hence the most sensitive test possible. For purposes of this study, the composite reading score is a combination of all subtest scores in reading on the Stanford Achievement Test which includes Word Reading, Paragraph Meaning, and Word Study Skills. The procedure used to obtain the composite reading score had been previously described in Chapter II. The Vocabulary subtest, which is a measure of a pupil's vocabulary independent of his reading ability and the Spelling subtest of the Stanford Achievement Test were not used to determine the composite reading score, but comparisons among the three treatment groups were made on each of these measures. To determine whether a linear relationship existed between IQ scores and the composite reading scores, a bivariate distribution was prepared for each of the three treatment groups. As can be seen by looking at these distributions, in Figures 2, 3, and 4, there is a very definite linear relationship between the two. Since linearity was observed in each distribution, the composite reading scores were then subjected to an analysis of variance for the three treatment groups by three IQ categories. Breakdown by sex was obtained for each of the subtests of the Stanford Achievement Test, but not on the composite reading score.

Table 29 illustrates the resulting means on the composite reading score.

TABLE 29

TABLE OF MEANS FOR THE THREE TREATMENT GROUPS ON COMPOSITE READING SCORE

		Lippincott N=189	i/t/a N=238	Scott, Foresman N=229
	High	11.33	11.03	8.47
IQ	Average	9.68	9.32	7.57
	Low	8.06	8.30	6.14

COMPOSITE READING SCORE

PINTNER-CUNNINGHAM IQ SCORE	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
137-140											1					1
133-136								1			1					2
129-132					1					1	2	1				5
125-128								1	1		4					6
121-124										3	1					4
117-120							2		2	4	3					11
113-116				1	1	1	3	1	3	4	5	1				20
109-112					2		2	2	3	4	5					18
105-108				2	1	1	1	4	2	6	4					21
101-104			1	1	2	1	3	3	3	5	6					25
97-100			2	2	2	3	3	2	1	2	3					20
93-96				2	2	3	3	2	1	3						16
89-92			2	2	1	3	1	2	1	1	1					14
85-88	1	1		1	2	2	1	2	2							12
81-84						3	2	1								6
77-80		2	1		1		1									5
73-76							1	1								2
69-72					1											1
65-68		1														1
61-64																0
Total	1	4	6	11	16	17	23	22	19	33	36	2				190

Figure 2. Bivariate distribution of composite reading scores and IQ scores of the Lippincott population.

COMPOSITE READING SCORE

PINTNER-CUNNINGHAM IQ SCORE

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
137-140											1					1
133-136																0
129-132								1		4	1					6
125-128				1				1		1	5					8
121-124						1	2		1	1	2					7
117-120							1	4	1	2	5	1				14
113-116					1	2	2	2	1	5	4	1				18
109-112				1	1	1	4	3	5	6	5					26
105-108				1	2	2	8	2	6	6	3					30
101-104				1	3	3	2	3	3	5	2					22
97-100		2	1	3	5	2	3	3	5	2	2					28
93-96		2	2	6	3	1	1	6	3	3						27
89-92			1		3	3	3	1	4	2	2					19
85-88		1	2	4	2	2		1	1	2						15
81-84				3		2				1						6
77-80	1		2	2	3						1					9
73-76		1			1			1								3
69-72																0
65-68																0
61-64					1											1
Total	1	6	8	22	25	19	26	28	30	40	33	2				240

Figure 3. Bivariate distribution of composite reading scores and IQ scores of the i/t/a population.

COMPOSITE READING SCORE

PINTNER-CUNNINGHAM IQ SCORE	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
137-140							1	1								2
133-136																0
129-132																0
125-128						2	1		2	1						6
121-124						4		1	1							6
117-120					1	1			1	3						6
113-116		1		6	4	2	1			1						15
109-112		1		3	2	1	1	1	3		1					13
105-108		1	4	4	1	4	3	2	2	1	1					23
101-104			2	3	2	2	1	3	3	2						18
97-100		2	6	3	5	4	4	2	2	1						29
93-96		5	9	9	2	5	3	2	1							36
89-92	1	2	6	4	7	2										22
85-88		1	5	6	4	5	3		2							26
81-84		2	4	1	2	1	1									11
77-80		1			2											3
73-76	1	1	4	1	1											8
69-72	1		1													2
65-68	1	1				1										3
61-64																0
Total	4	18	41	40	33	34	19	12	17	9	2					229

Figure 4. Bivariate distribution of composite reading scores and IQ scores of the Scott, Foresman population.

As can be seen in Table 29, children of high intelligence, utilizing the Lippincott and Early-to-Read i/t/a series had approximately equivalent means. The mean composite reading score for the high IQ Scott, Foresman group was lower than the means of the other two treatment groups. A similar difference in mean scores was obtained for the average and low IQ children in each of the three treatment groups. In each case, Lippincott and Early-to-Read i/t/a were equivalent, whereas Scott, Foresman was lower.

The significance of the observed differences can be seen in Table 30 which illustrates the analysis of variance computed among the three treatment groups by IQ category (high, middle, and low) for the composite reading score.

TABLE 30

ANALYSIS OF VARIANCE TABLE FOR COMPOSITE READING SCORE

Sources	df	Mean Square	F
Treatments (T)	2	4.97	6.71**
IQ	2	5.78	
T x IQ	4	0.09	1.15
Error	650	.0745	

**Significant at the .01 level of confidence.

From Table 30, it can be seen that the observed differences in means among the three experimental groups were significant at the .01 level for 2 and 650 degrees of freedom. Since the mean differences between the Lippincott and Early-to-Read i/t/a groups were extremely small, it may be inferred that the superiority of these two basal systems over the Scott, Foresman system produced this significant result. The F-ratio among the three basal reading systems was computed as 6.71. No significant interactions between treatments and IQ occurred. Hence, it may be inferred that intelligence did not have differential effects among the three basal reader systems, but, in fact, intelligence produced a similar effect regardless of the basal system employed.

Word reading

The Word Reading subtest of the Stanford Achievement Test, Primary I Battery, measures the ability of a pupil to analyze or recognize a word without contextual clues. The test consists of 35 multiple-choice items in which the pupil is required to look at a picture and then select the word which stands for the picture from a group of four words.

Table 31 gives the mean scores for each of the three treatment groups by intelligence category and by sex.

TABLE 31

TABLE OF MEANS FOR THE THREE TREATMENT GROUPS ON WORD READING SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Group Variables		Lippincott N=189		i/t/a N=238		Scott, Foresman N=229		Summary
Sex		Boys	Girls	Boys	Girls	Boys	Girls	
IQ	High	29.19	31.24	28.44	29.02	17.50	21.94	26.22
	Average	26.81	27.20	24.25	24.42	19.5	19.67	23.65
	Low	22.38	19.57	21.04	23.41	14.39	15.14	19.32
	Summary	26.13	26.003	24.58	25.62	17.14	18.92	23.06
		26.07		25.10		18.03		

In Table 31, it may be seen that the mean scores obtained by the Lippincott and _Early-to-Read i/t/a_ groups were approximately equivalent, while the Scott, Foresman group had a considerably lower mean score in _Word Reading_. These differences were maintained for each of the three categories of intelligence and also by sex. It is also interesting to observe that the high IQ children learning to read in the Scott, Foresman series had a lower mean score than the average IQ children learning in that same basal series. In addition, the high IQ Scott, Foresman group had a lower mean score in _Word Reading_ than the average IQ children in the Lippincott and _Early-to-Read i/t/a_ groups. The high IQ Scott, Foresman group scored no better than the low IQ children in the other two treatment groups. The differences by sex illustrate that girls had a higher mean score for each of the intelligence categories in each of the basal reader systems. The comparisons among the three basal reader systems also suggest small differences in means between the Lippincott and _Early-to-Read i/t/a_ series and a much lower mean for the Scott, Foresman series.

An analysis of variance was computed between the three basal reader systems for high, middle, and low IQ children and by sex to determine whether the differences observed in Table 21 were significant or could have occurred as a result of chance. The summary of the analysis of variance is presented in Table 22. As can be seen from Table 22, the F-ratio computed among the three basal reader systems was an astronomically high 50.67 for 2 and 638 degrees of freedom. This difference would be significant at beyond the .01 level. From an analysis of the means reported in Table 31, it may be inferred that the significance of this difference was the result of the lower mean scores obtained on _Word_

TABLE 32

ANALYSIS OF VARIANCE TABLE FOR
WORD READING SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Sources	df	Mean Square	F
Treatments (T)	2	115.53	50.67**
Sex	1	3.61	1.58
IQ	2	72.93	
T x Sex	2	1.38	<1
T x IQ	4	4.53	1.99
Sex x IQ	2	2.40	1.05
T x Sex x IQ	4	2.02	<1
Error	638	2.28	

**Significant at .01 level of confidence

Reading by the Scott, Foresman group. Despite the fact that the girls had a higher mean score in each of the three basal reader systems for each of the three levels of intelligence, the differences observed could have been due to chance. The fact that each of the observed difference favors the girls suggests that further study in this area is necessary. The interaction between basal reader systems and sex was less than one on the computed F-ratio, which suggests that sex produced no differential effects among the three basal reader systems, but instead, had the same effect in each.

The graph in Figure 5 illustrates that there is a linear relationship among the basal reader systems and IQ subclass and that the lines among each of the three systems are relatively parallel. This illustrates that no significant interaction between IQ and treatment occurred. The F-ratio computed for this interaction was 1.05, which does not approach significance. Hence, intelligence had the same effect on each of the three basal reader systems, with no differential effects occurring.

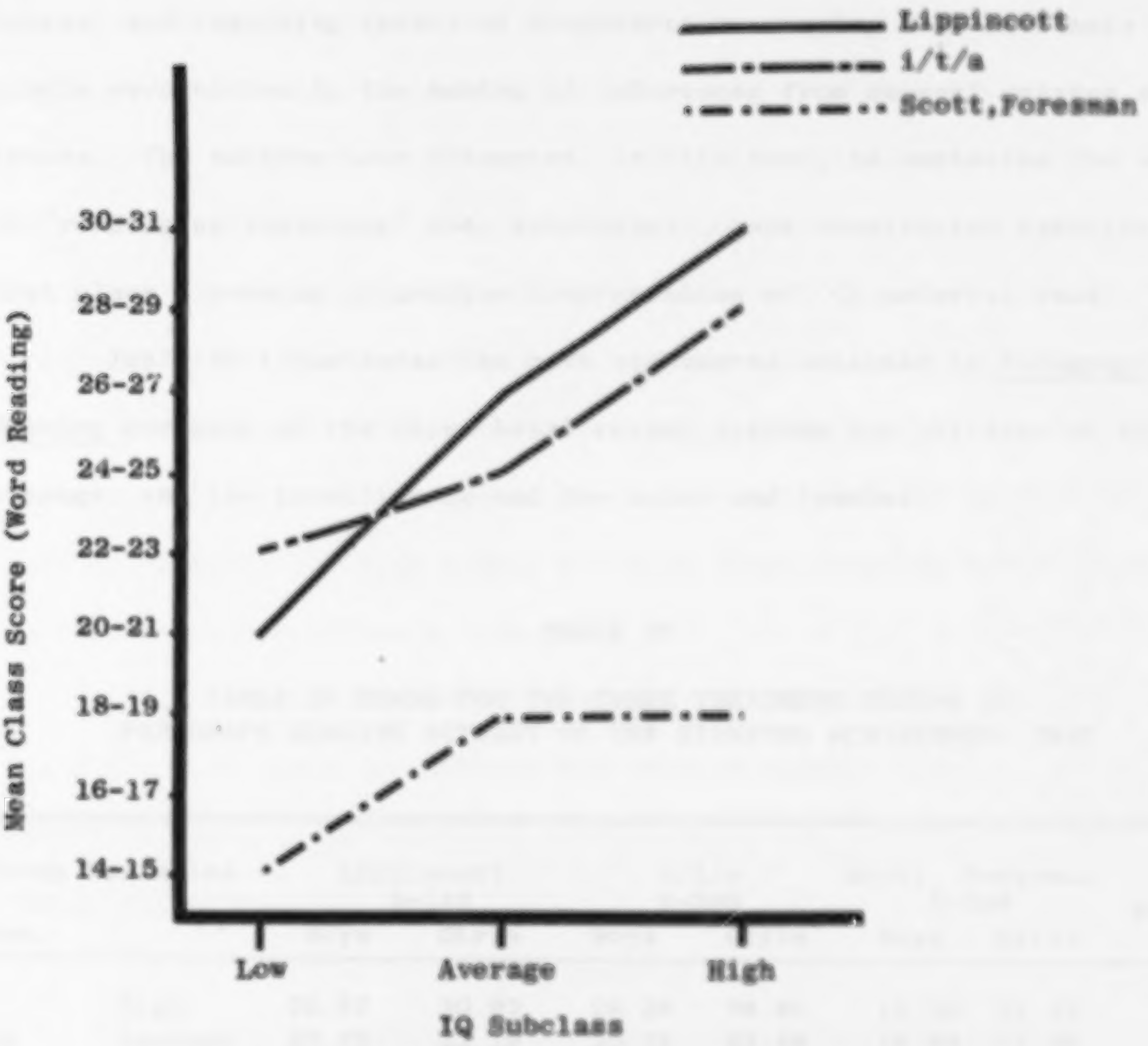

Figure 5 Relationship between treatment and IQ on Stanford Word Reading.

Paragraph Meaning

This subtest of the Stanford Achievement Test consists of a series of paragraphs, graduated in difficulty, from each of which one or more words have been omitted. The pupil's task is to demonstrate his comprehension of the paragraph by selecting the proper word for each omission from four choices which are afforded him. The test provides a functional measure of the child's ability to comprehend connected discourse ranging in length from single sentences to paragraphs of six sentences, and involving levels of comprehension varying from extremely simple recognition to the making of inferences from several related sentences. The authors have attempted, in this test, to emphasize the notion of "reading as reasoning" and, accordingly, have constructed exercises that place a premium on genuine comprehension of the material read.

Table 33 illustrates the mean raw scores obtained in Paragraph Meaning for each of the three basal reader systems for children of high, average, and low intelligence and for males and females.

TABLE 33

TABLE OF MEANS FOR THE THREE TREATMENT GROUPS ON PARAGRAPH MEANING SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Group Variables		Lippincott N=189		i/t/a N=238		Scott, Foresman N=229		Summary
Sex		Boys	Girls	Boys	Girls	Boys	Girls	
IQ	High	26.67	30.92	28.28	28.80	15.00	22.49	25.36
	Average	20.62	23.10	20.75	22.58	16.69	17.98	20.29
	Low	17.21	16.43	17.42	20.67	12.33	13.68	16.29
		21.50	23.50	22.20	24.01	14.70	18.04	
	Summary							20.64
		22.50		23.10		16.40		

Examination of Table 33 reveals that the Lippincott and Early-to-Read i/t/a groups obtained means that were relatively equivalent for each of the three categories of intelligence and also between sex. The Scott, Foresman group, again, obtained lower means for each of the three categories of intelligence and two sexes than the other two treatment groups. On this subtest, the high IQ Scott, Foresman boys had a lower mean score than those boys classified as being of average intelligence. Scott, Foresman girls, however, categorized as being of high intelligence, had a higher mean score than the average IQ girls. The high IQ Scott, Foresman boys did no better than the low IQ boys who were taught by the Lippincott or Early-to-Read i/t/a series. The average IQ Scott, Foresman boys and girls obtained mean scores that were lower than the average IQ of Lippincott and Early-to-Read i/t/a boys and average IQ of Lippincott and Early-to-Read i/t/a girls respectively. Again, on this subtest, the mean score obtained by the girls for each IQ category was higher than the mean score obtained by the boys in each of the three treatment groups.

To determine whether the differences in means were significant, a 3 x 3 x 2 analysis of variance was used to analyze the three basal reader systems, three levels of intelligence, and two sexes. Table 34 illustrates these results:

TABLE 34

ANALYSIS OF VARIANCE TABLE FOR PARAGRAPH MEANING SUBTEST
OF THE STANFORD ACHIEVEMENT TEST

Sources	df	Mean Square	F
Treatments (T)	2	83.16	15.93**
Sex	1	26.08	5.00*
IQ	2	124.03	
T x Sex	2	1.05	< 1
T x IQ	4	7.16	1.37
Sex x IQ	2	3.30	< 1
T x Sex x IQ	4	3.61	< 1
Error	638	5.22	

*Significant at the .05 level of confidence
**Significant at the .01 level of confidence

The F-ratio computed on the three basal reader systems was 15.93. This would be significant at beyond the .01 level for 2 and 638 degrees of freedom. Hence, the differences observed in the table of means were very significant, and suggest that the Lippincott and Early-to-Read i/t/a series produced significantly better results in Paragraph Meaning. The differences between sexes in which the girls had a higher mean score for each category of intelligence than the boys in each of the three treatment groups were also significant at the .01 level for 1 and 638 degrees of freedom. The computed F-ratio was 5.00. Hence, one may infer that the achievement of girls is significantly higher than the achievement of boys regardless of the basal reader system used. The interactions of the three basal reader systems and sex resulted in a computed F-ratio of less than one, which suggests that sex does not have a differential effect in any of the treatments, but the girls do better than the boys regardless of the system used.

The following graph in Figure 6 illustrates the effect of the basal reader systems by IQ subclass. Again, a strong linear relationship can be observed. The fairly similar slopes of the lines suggest no interaction between intelligence and the basal reader system. This was verified by computing the F-ratio for this interaction, which resulted in a ratio of less than one. Hence, intelligence produces similar effects in each of the three basal reader systems.

Word Study Skills

This subtest of the Stanford Achievement Test consists of fifty-six multiple-choice items broken down to measure the auditory perception of beginning and ending sounds, phonics, and phonograms. In the first

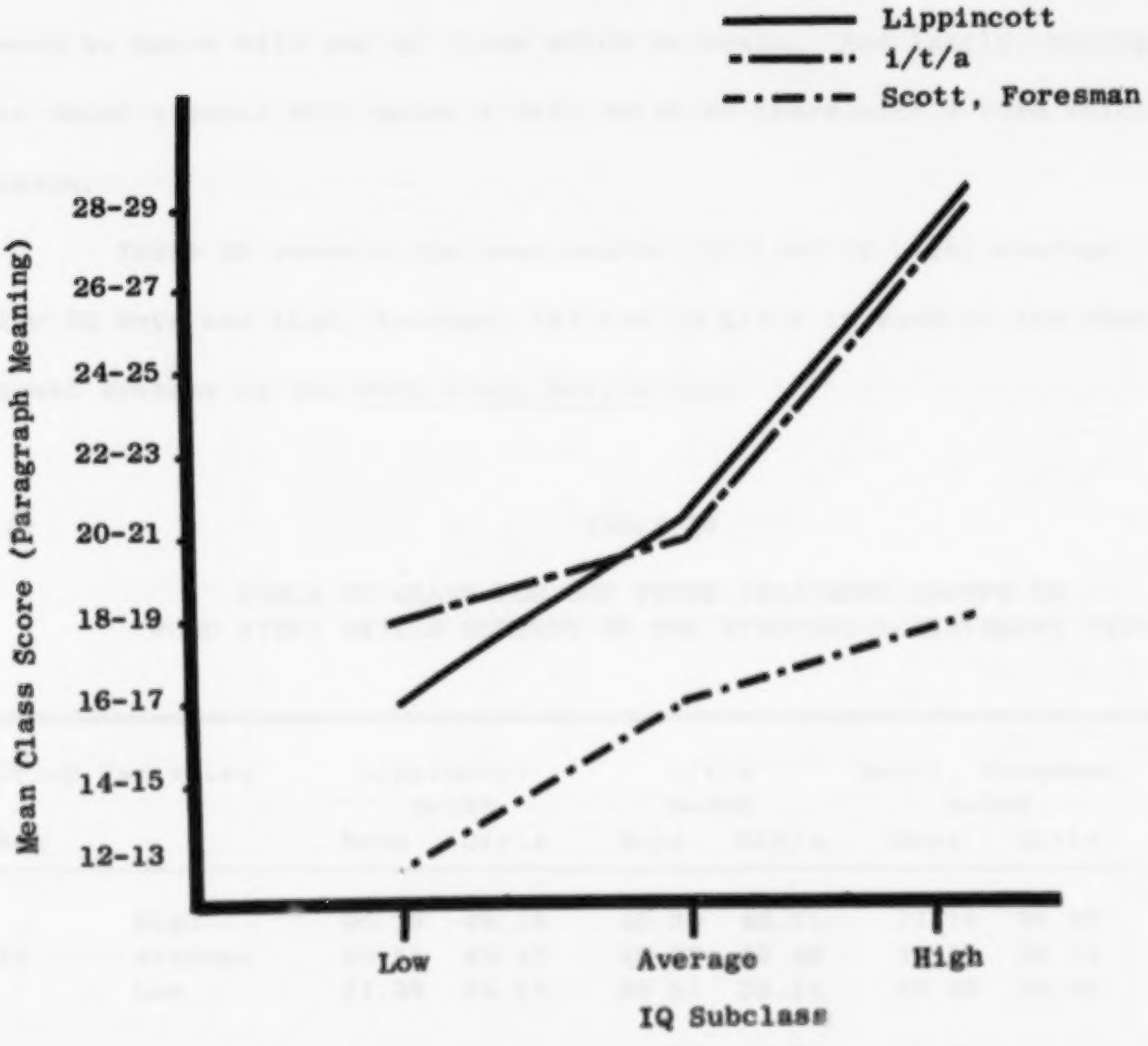

Figure 6 Relationship between treatment and IQ on Stanford Paragraph Meaning.

part of this test, a pupil hears one word read by the teacher and then reads silently as the teacher reads aloud three other words from which he selects one containing a beginning sound which is the same as the beginning sound in the word the teacher read. Next is auditory perception of ending sounds in which the word to be chosen has the same ending sound as the word which the pupil hears. In the third part of this test, the pupil selects the written word which is the same as the last word in the sentence read by the teacher. The pupil matches the word he hears with one of three which he reads. And lastly, phonograms, in which a pupil must match a word which he hears with a word which he reads.

Table 35 reveals the mean scores obtained by high, average, and low IQ boys and high, average, and low IQ girls in each of the three basal systems on the Word Study Skills test.

TABLE 35

TABLE OF MEANS FOR THE THREE TREATMENT GROUPS ON WORD STUDY SKILLS SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Group Variables		Lippincott N=189		i/t/a N=238		Scott, Foresman N=229		Summary
Sex		Boys	Girls	Boys	Girls	Boys	Girls	
	High	46.28	49.18	45.50	46.71	37.18	39.86	44.12
IQ	Average	40.51	42.43	40.29	42.48	35.31	36.10	39.52
	Low	37.38	33.57	36.51	39.11	29.48	30.95	34.50
		41.39	41.73	40.77	42.77	33.99	35.64	
	Summary							39.38
		41.56		41.77		34.81		

From Table 35 it may be observed that the mean scores for boys in each of the IQ categories were approximately equivalent for the Lippincott and Early-to-Read i/t/a groups. The mean scores obtained for boys by level of intelligence in the Scott, Foresman group were lower than those obtained in the other two treatment groups. The girls, in each case, for each of the three categories of intelligence, obtained a higher mean score than the boys within each of the three treatment groups, with only one exception: the low IQ boys in the Lippincott group had a higher mean score than the low IQ girls in that group. In the Early-to-Read i/t/a group, low IQ girls had a higher mean score than the low IQ girls in the Lippincott group.

To determine whether any of the observed differences in the means were significant, an analysis of variance was computed among the three basal reader systems by IQ category (high, average, and low), and sex.

TABLE 36

ANALYSIS OF VARIANCE TABLE FOR WORD STUDY SKILLS SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Sources	df	Mean Square	F
Treatments (T)	2	93.92	22.31**
Sex	1	7.93	1.88
IQ	2	138.83	
T x Sex	2	1.15	< 1
T x IQ	4	3.27	< 1
Sex x IQ	2	1.89	< 1
T x Sex x IQ	4	2.70	< 1
Error	638	4.21	

**Significant at the .01 level of confidence.

As can be seen in Table 36, the F-ratio computed among the three basal reader systems for 2 and 638 degrees of freedom was 22.31, which is significant at well beyond the .01 level of confidence. Virtually no differences were observed in the means of the Lippincott and Early-to-Read i/t/a groups in each category of intelligence. It may be inferred that this result is due to the lower mean scores for the Scott, Foresman group in each of the three categories of intelligence for both boys and girls.

Despite the fact that the girls had higher mean scores than the boys in almost all cases, with one exception, the F-ratio computed for the category of sex was 1.88, which is not significant. The F-ratio computed for the interaction of basal reader systems and sex was less than one, which suggests that the few differences occurring in the area of sex were due to chance, and that sex does not produce differential effects in each of the three basal reader systems, but has similar effects in each. The interaction between treatments, sex, and IQ was also not significant. The F-ratio computed was less than one.

The graph in Figure 7 illustrates the effects of basal reader systems by IQ subclass.

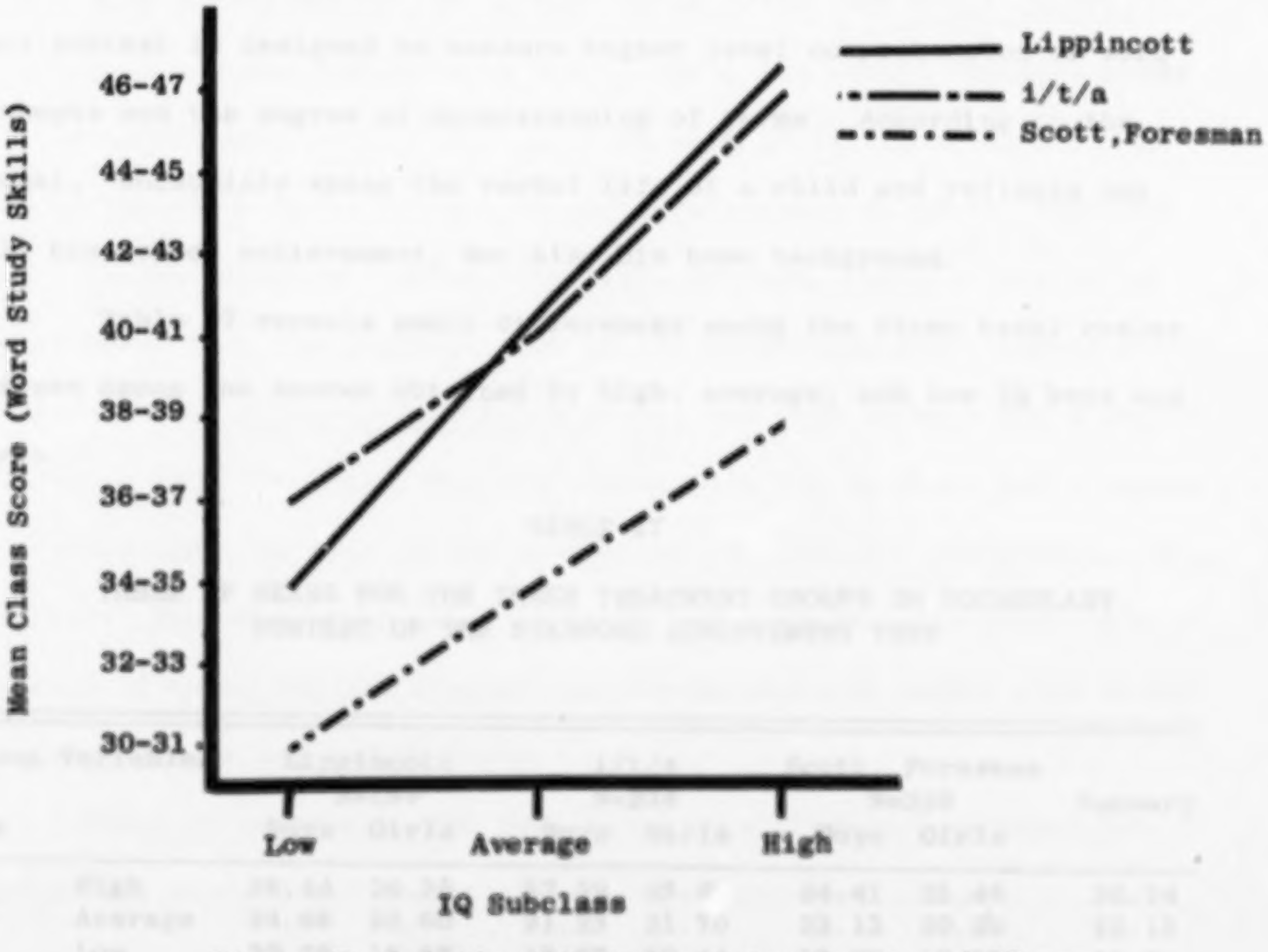

Figure 7 Relationship between treatment and IQ on Stanford Word Study Skills.

As can be seen from Figure 7, a strong linear relationship between intelligence and achievement in each of the three basal reader systems is illustrated. Slopes of the lines are relatively similar, which suggests no interaction between IQ and basal reader system. The F-ratio computed for this interaction was less than one, which for 4 and 638 degrees of freedom is, certainly, not significant.

Vocabulary

In the Vocabulary subtest of the Stanford Achievement Test, the pupil is required to select from a series of three alternatives the correct answer to a question or statement read by the teacher. These items measure knowledge of synonyms, simple definitions, and ready associations. This subtest is designed to measure higher level comprehension of word concepts and the degree of understanding of terms. According to the manual, "vocabulary spans the verbal life of a child and reflects not only his school achievement, but also his home background."

Table 37 reveals small differences among the three basal reader systems among the scores obtained by high, average, and low IQ boys and girls.

TABLE 37

TABLE OF MEANS FOR THE THREE TREATMENT GROUPS ON VOCABULARY SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Group Variables		Lippincott N=189		i/t/a N=238		Scott, Foresman N=229		Summary
Sex		Boys	Girls	Boys	Girls	Boys	Girls	
IQ	High	28.44	26.39	27.29	25.45	24.41	25.46	26.24
	Average	24.68	22.80	21.23	21.70	22.13	20.26	22.13
	Low	22.79	18.57	18.67	20.44	18.00	16.95	19.24
	Summary	25.31	22.59	22.40	22.53	21.51	20.89	
		23.95		22.46		21.20		22.54

As can be seen in Table 37, the high IQ boys from the Lippincott group had the highest mean score, while the Scott, Foresman boys in the high IQ category had the lowest mean score. Among the average IQ boys, the Lippincott boys had the highest score, and the Early-to-Read i/t/a boys had the lowest score. In the low IQ category, the Lippincott boys had the highest score, and the Scott, Foresman boys the lowest. Among the girls, the high IQ girls in the Lippincott group had the highest mean score in Vocabulary, while the Early-to-Read i/t/a girls and Scott, Foresman girls were virtually identical and were lower. In the average IQ category, the Lippincott girls had the highest mean score, the Scott, Foresman girls the lowest score, but the differences between the highest and lowest scores were fairly small. Among the low IQ girls, the Early-to-Read i/t/a girls had the highest mean score, while the Scott, Foresman girls had the lowest mean score. The girls in each of the IQ categories for each of the three basal systems do not consistently do better than the boys. In the Lippincott group, the high, average, and low IQ boys had higher means than the girls in each of the respective intelligence categories. In the Early-to-Read i/t/a group, the high IQ boys had a higher mean score than the girls, but the average and low IQ girls had a higher mean score than the boys in the respective intelligence categories. In the Scott, Foresman group, high IQ girls had a higher mean score than the high IQ boys, but the average and low IQ boys had higher mean scores than the average and low IQ girls.

Table 38 represents the results of an analysis of variance for each of the three basal reader systems by IQ category and sex.

TABLE 38

ANALYSIS OF VARIANCE TABLE FOR VOCABULARY SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Sources	df	Mean Square	F
Treatments (T)	2	11.35	6.84**
Sex	1	5.13	3.09
IQ	2	74.38	
T x Sex	2	3.27	1.97
T x IQ	4	.49	<1
Sex x IQ	2	0.02	<1
T x Sex x IQ	4	1.81	1.09
Error	638	1.66	

**Significant at the .01 level of confidence.

As indicated in Table 38, although the differences among the three treatments by IQ and sex were small, the F-ratio computed for 2 and 638 degrees of freedom was 6.84 which was significant at the .01 level. Differences between the sexes were not significant at the .05 level for 1 and 638 degrees of freedom, as the F-ratio was 3.09. The interactions on this subtest were not significant for either sex or IQ by basal reader system.

Spelling

The _Spelling_ subtest of the _Stanford Achievement Test_, is measured by a dictation type exercise in which the word to be spelled is pronounced by the teacher, a sentence is read, and the word is repeated, following which the pupil writes the word as he would normally do in his own writing. On the _Spelling_ subtest, those children who were taught to read in the _Early-to-Read i/t/a_ series, were scored on the basis of whether the word was spelled correctly by the standards of traditional orthography. If a child spelled a word correctly in i/t/a, but this spelling varied from the correct spelling of the word in traditional orthography, his response was considered to be incorrect.

Table 39 illustrates the performance of high, average, and low IQ boys and girls taught by each of the three basal reader systems.

TABLE 39

TABLE OF MEANS FOR THE THREE TREATMENT GROUPS ON SPELLING SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Group Variables		Lippincott N=189		i/t/a N=238		Scott, Foresman N=229		Summary
Sex		Boys	Girls	Boys	Girls	Boys	Girls	
	High	16.26	17.64	12.21	14.60	11.14	13.66	14.52
IQ	Average	13.46	14.20	10.59	12.34	10.91	11.90	12.22
	Low	11.59	9.57	9.47	10.37	7.48	8.97	9.58
		13.60	13.67	11.12	12.47	9.84	11.51	
	Summary							12.10
		13.64		11.80		10.68		

An inspection of Table 39 reveals that those children taught to read in the Lippincott basal series had higher mean scores than those children who were taught in the *Early-to-Read i/t/a* or the Scott, Foresman series. These differences, with Lippincott being superior, were true in each of the three IQ categories for boys and in each of the three IQ categories for girls, with the exception of the low IQ *Early-to-Read i/t/a* girls whose score was higher than the low IQ girls in the Lippincott or Scott, Foresman series. In all categories, but one, the girls had a higher mean score on *Spelling* than did the boys, with the sole exception that the low IQ Lippincott boys scored higher than the low IQ Lippincott girls.

To determine whether any of the differences observed in mean scores were significant, an analysis of variance was computed among the three basal reader systems for high, average, and low IQ boys and for high, average, and low IQ girls. Table 40 reveals the results of the analysis of variance.

TABLE 40

ANALYSIS OF VARIANCE TABLE FOR SPELLING SUBTEST OF THE STANFORD ACHIEVEMENT TEST

Sources	df	Mean Square	F
Treatments (T)	2	15.42	10.21**
Sex	1	4.76	3.15*
IQ	2	36.70	
T x Sex	2	1.07	< 1
T x IQ	4	1.41	< 1
Sex x IQ	2	1.10	< 1
T x Sex x IQ	4	.49	< 1
Error	638	1.51	

*Significant at the .05 level of confidence.
**Significant at the .01 level of confidence.

Table 40 indicates that the differences observed among the three basal reader systems were significant at beyond the .01 level of confidence. Hence, for 2 and 638 degrees of freedom, the computed F-ratio among the basal reader systems was 10.21. The differences observed in means among the boys and girls on the Spelling subtest reveal a significant difference at the .05 level of confidence, as for 1 and 638 degrees of freedom. The computed F-ratio was 3.15. This indicates that girls do significantly better than boys in Spelling in each of the three basal reader systems. There was no significant interaction between basal reader systems and sex, or between sex and IQ category. The computed F-ratio for each of these interactions was less than one. This suggests that sex differences were the same regardless of the basal reader system employed.

The graph in Figure 8 illustrates the effect of each basal reader system by IQ subclass. As can be observed in Figure 8, there is a linear relationship between IQ and Spelling achievement in each of the three basal reader systems. The slopes of the lines in each of the three groups are not significantly different, as can be observed by computing the interaction between basal reader system and IQ. The computed F-ratio for 4 and 638 degrees of freedom was less than one, which

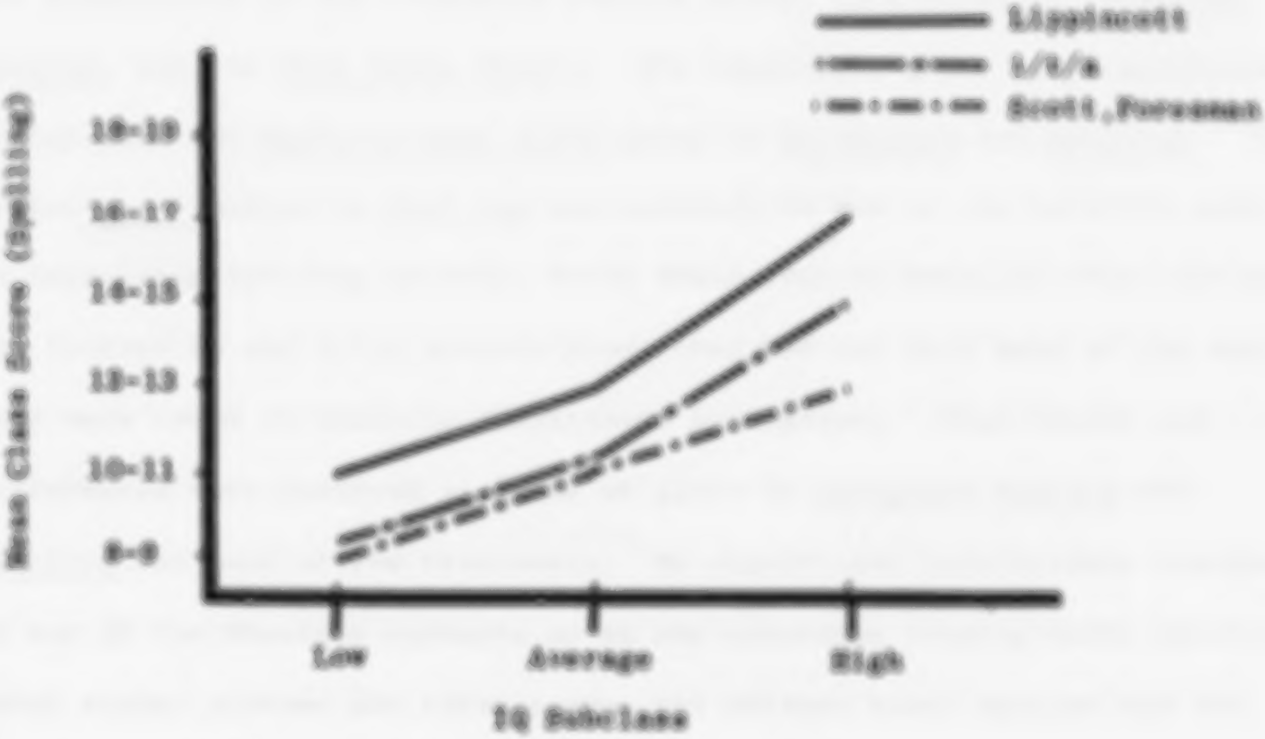

Figure 8 Relationship between treatment and IQ on Stanford Spelling.

infers that IQ has similar effects regardless of the basal reader system employed. Critical ratios computed between treatment groups indicated that Lippincott was significantly better than *Early-to-Read i/t/a* or Scott, Foresman groups in *Spelling*, and the i/t/a group was significantly better than the Scott, Foresman group in *Spelling*.

Summary statement

On the composite reading score and on each of the five subtests of the *Stanford Achievement Test* - *Word Reading*, *Paragraph Meaning*, *Word Study Skills*, *Vocabulary*, and *Spelling*, there were significant differences among the three treatment groups. In each analysis, the Scott, Foresman group was significantly lower than the i/t/a or Lippincott groups. Differences between the Lippincott and the *Early-to-Read i/t/a* groups were not significant on the composite reading score, *Word Reading*, *Paragraph Meaning*, and the *Word Study Skills*. The Lippincott group was significantly better than the *Early-to-Read i/t/a* group in *Vocabulary* and *Spelling*. The difference obtained in *Spelling* may possibly be due to the criteria used in appraising spelling ability, which would tend to penalize those children who learned in the i/t/a series, since they had not seen many of the words they were asked to spell in traditional orthography. Significant sex differences were observed in favor of girls in *Paragraph Meaning* and *Spelling* for each of the treatments. No significant interactions occurred on any of the Stanford subtests or on the composite reading score between basal reader systems and intelligence and between basal systems and sex.

Use of Readiness Tests to Predict Reading Achievement

Since significant differences were observed among treatments, the reading readiness tests were examined to determine which of these tests correlated best with the criteria of reading achievement, and whether the best predictions were obtained by different readiness tests for each of the basal systems.

Canonical correlations

To determine whether it was possible to differentially predict the best chances of success in each of the three basal reader systems, a canonical correlation (Hotelling, 1935) was computed. In the canonical correlation, 14 predictors or 15 predictors were compared to three criteria. The predictors used were the following reading readiness tests: Murphy-Durrell Phonemes; Letter Names: capitals, lower-case, and total capital and lower-case letters, and, Learning Rate; Thurstone Pattern Copying, and Identical Forms; Metropolitan Word Meaning, Listening, Matching, Numbers, Copying, Alphabet, and total score. A fifteenth predicting variable, the IQ score of the Pintner-Cunningham Primary Test was also added. The three criteria utilized were the three reading subtests of the Stanford Achievement Test, Primary I, which were Word Reading, Paragraph Meaning, and Word Study Skills. The correlation matrices for the three treatment groups are contained in Appendices D through F, page 121 to page 123. The correlation matrices contain the intercorrelations among each of the fifteen predictors and the three criteria. The computer was programmed to selectively drop predicting variables to determine which combination of predictors would result in the highest multiple correlation with the criteria. Three variables were found to give the best prediction with the criteria, so that the addition of the other twelve predictors did not significantly improve the multiple correlation. The three predictors thus determined were (a) Murphy-Durrell Phonemes, (b) Murphy-Durrell Letter Names (total capital and lower-case letters), and (c) the total score of the Metropolitan Readiness Tests. A fourth predictor was the IQ score of the Pintner-Cunningham Primary Test, added to each of the predicting variables to determine whether any significant improvement in prediction could be obtained. Table 41 reveals the results of the canonical correlations with IQ added as a predictor and without IQ added as a predictor for boys and girls separately in each of the three treatment groups.

TABLE 41

CANONICAL CORRELATIONS BY SEX AND TREATMENT

PREDICTORS: MURPHY-DURRELL PHONEMES, MURPHY-DURRELL LETTER NAMES, METROPOLITAN TOTAL SCORE, PINTNER-CUNNINGHAM IQ
READING ACHIEVEMENT CRITERIA: STANFORD ACHIEVEMENT WORD READING, WORD STUDY SKILLS AND PARAGRAPH MEANING

	N	With IQ	Without IQ
Lippincott			
Boys	107	.76	.76
Girls	82	.83	.83
i/t/a			
Boys	123	.74	.73
Girls	115	.78	.77
Scott, Foresman			
Boys	118	.71	.71
Girls	111	.77	.77

As can be seen in Table 41, the prediction of success for boys or girls in each of the three basal reader systems did not improve significantly by including IQ as one of the predicting variables. By utilizing a z transformation and computing the standard error of the difference, it was found that no significant difference resulted between the prediction of reading achievement with and without IQ as a predicting variable. Also, by utilizing a z transformation and computing the standard error of the difference, it was found that no significant differences in prediction between boys and girls occurred for any of the three basal reader systems. Hence, according to the obtained results, no differential prediction is possible to determine in advance which of the three basal reader systems would result in the greatest chances of reading success. The same predicting variables resulted in the highest correlations with the criteria in each of the three treatment groups.

Partial correlations

As can be seen from the correlation matrices, contained in Appendices D through F, page 121 to page 123 intelligence is significantly related to each of the readiness variables used as predictors in this study. Hence, the canonical correlations were computed with IQ as a predictor and without IQ as a predictor to determine whether any significant difference in the ability to predict would be discovered. Despite the fact that significant differences did not occur when intelligence was or was not used as a predicting variable, partial correlations (Guilford, 1965, pp. 339-341) were computed to determine whether the best predicting variables of reading achievement would still be significant if intelligence were partialled out. The best predictors discovered in the canonical correlations were *Murphy-Durrell Phonemes*, *Murphy-Durrell* total capitals and lower-case *Letter Names* subtest, and the total score of the *Metropolitan Readiness Tests*. Partial correlations were, thus, computed between each of these predicting variables with each of the criteria variables which included the following: *Stanford Word Reading*, *Paragraph Meaning*, and *Word Study Skills*. Each of the major predictors were correlated with each of the criteria with intelligence partialled out for each of the three treatment groups, for boys and girls separately, and for the total groups. In each case, the resulting partial correlations were significant at the .01 level of confidence for 78 of the 81 partial correlations computed. Of the three partial correlations that were not significant at the .01 level, two were significant at the .05 level, and one was not significant at either level of confidence. Hence, the readiness tests employed as predictors are measuring other factors associated with reading success other than intelligence. The use of these predictors in computing canonical correlations is, therefore, justifiable. The computed partial correlations are contained in Appendices G through I, page 124 to page 126. As can be seen from the canonical correlations, the prediction of reading success is possible, utilizing the most predictive subtests in the readiness battery, and that for this purpose, only three of the subtests utilized in the readiness battery need be utilized. Unfortunately, it is not possible to differentially predict for individual children which of the basal reader systems would yield their best chances of success. Hence, the decision as to which basal reader system should be utilized can be determined only by which system produced the best reading

achievement. In this study, the best systems proved to be the Lippincott and *Early-to-Read i/t/a* procedures, and the achievement of the children taught in these two systems was approximately equivalent.

CHAPTER 4

SUMMARY AND CONCLUSIONS

The basic purpose of this study was to determine the effects of three different basal reader systems containing discrete features on the reading achievement of first-grade children of different levels of intelligence and sex. The three basal systems were the Lippincott Basic Reading series, the Early-to-Read i/t/a program, and the Scott, Foresman New Basic Readers series. Approximately 630 children in twenty-six classes were included in the sample which was drawn from three school districts in Long Island, New York. The procedures utilized to conduct the experiment have been described in earlier chapters of this report. To determine whether any differential effects occurred in reading achievement, the interactional effects of the three experimental treatments, three levels of intelligence, and the two sexes were analyzed. Since significant differences were observed among treatments, the reading readiness tests were examined to determine which combination of tests correlated best with the criteria of reading achievement. In addition, an attempt was made to determine whether the best predictions were obtained by different readiness tests for each of the basal reader systems.

The conclusions drawn from this study and interpretation of their possible implications for first-grade reading instruction follow:

1. There was a significant difference on a test of reading and spelling achievement among children taught to read by three different basal reader systems, containing discrete features, at the end of first-grade. The three basal reader systems utilized were the Lippincott

Basic Reading series, the Early-to-Read i/t/a series, and the Scott, Foresman New Basic Readers series. The Lippincott and Early-to-Read i/t/a groups scored significantly higher than the Scott, Foresman group on the composite reading score, which was a composite of the Word Reading, Paragraph Meaning, and Word Study Skills subtests of the Stanford Achievement Test, Primary I and on all of the subtests -- Word Reading, Paragraph Meaning, Word Study Skills, and Spelling -- of the Stanford Achievement Test.

2. The Lippincott series produced significantly higher achievement on the subtests of Vocabulary and Spelling of the Stanford Achievement Test than the Early-to-Read i/t/a group

3. Sex was not a factor producing differential results among any of the three basal reader systems. That is, no one basal reader system seemed to produce better results with boys than with girls, or vice versa. In each of the three basal reader systems, the girls achieved higher mean scores than the boys. Significant differences between girls and boys occurred on the Spelling and Paragraph Meaning subtests. This result suggests the possibility that in first-grade boys and girls should be instructed separately regardless of the basal reader system used. It also suggests that a greater delay in teaching boys to read may be useful. These are tenable hypotheses which can be inferred from these results. Further experimental verification is necessary to prove or disprove these hypotheses.

4. In each of the three basal reader systems, intelligence was not a major factor in distinguishing probable chances for success. No one of the basal reader systems proved to be better at any IQ range than the others. Generally, the same results occurred. The Lippincott and

i/t/a series produced the best results regardless of IQ, and the Scott, Foresman series produced the lowest results regardless of IQ. One interesting variation occurred. Children of average intelligence achieved higher mean scores than those of high intelligence on the *Word Reading* and *Paragraph Meaning* subtests of the *Stanford Achievement Test, Primary I*. These differences, however, were not significant. Hence, the Lippincott and *Early-to-Read i/t/a* series produced better achievement at each of the IQ ranges on each of the *Stanford* subtests than did the Scott, Foresman basal series. Also, children of high intelligence, in the Scott, Foresman group, tended to score significantly lower than children of average intelligence in the Lippincott and *Early-to-Read i/t/a* groups. This was true despite the fact that all of the children in the average IQ range in the Lippincott and *Early-to-Read i/t/a* groups were lower in IQ than all of the children in the high IQ range in the Scott, Foresman group.

3. The pattern of results was very consistent on the post-instruction oral reading and word recognition measures administered to very small randomly selected subsamples of the three treatment groups. The results indicated that the Lippincott and i/t/a pupils were comparable on all of the measures, and that both groups were significantly superior to the Scott, Foresman pupils on the *Gates Word Pronunciation Test*, *Karlsen Phonemic Word Test*, and *Phonetically Regular Words Oral Reading Test*. There was no significant difference between the i/t/a and Scott, Foresman pupils on *Gilmore Rate*, but the i/t/a pupils scored significantly higher in *Accuracy*. The reverse of the aforementioned results was observed between the Lippincott and Scott, Foresman pupils;

Lippincott pupils scored significantly higher in Rate than Scott, Foresman pupils but no significant difference between the groups existed in Accuracy. The observed differences in results suggest that, at the end of grade one, the i/t/a and Lippincott basal reader series produce significantly better results than the Scott, Foresman series in word recognition skills and oral reading ability.

6. The same readiness factors produced the highest prediction with the criterion of reading achievement for each of the three treatment groups. Hence, it is not possible to differentially predict which children have a greater chance of reading success in advance of their being instructed in each of the three basal reader systems utilized in this study.

7. The reading achievement of the Early-to-Read i/t/a group was not significantly different from the reading achievement of the Lippincott basal reader group, despite the fact that the Lippincott group had significantly higher readiness scores in almost all areas of the Metropolitan Readiness Test and the Murphy-Durrell Diagnostic Readiness Test than did the i/t/a children. Also, it should be noted that children instructed in i/t/a performed at a similar level to those children who were instructed in T.O. in the Lippincott series, although children taught with i/t/a were penalized because all the final measurements were taken using traditional orthography. The penalty, of course, was greatest for those children who had not as yet made the transition to traditional orthography. Long term effects of i/t/a would have to be observed before definitive conclusions could be made.

8. The Lippincott and Early-to-Read i/t/a basal reader programs

generally utilized an approach to reading that is considerably more analytic than the Scott, Foresman program. This is manifested in an instructional approach which places heavy emphasis upon word structure and the phonetic characteristics of words. In order to make full utilization of this approach, it is necessary to teach a large number of phonic skills in first-grade which would ordinarily be taught over a period of three years in a typical basal reader system like Scott, Foresman. Hence, those children taught to read in the Scott, Foresman basal series will eventually be taught the same skills developed in the Lippincott and *Early-to-Read i/t/a* readers, but many of these skills would not appear until second- or third-grade within the program. It is possible a stronger base in these skills is developed through the slower pace of introduction and the greater amount of repetition of these skills in the Scott, Foresman program. Hence, the significant differences in achievement in favor of the Lippincott and *Early-to-Read i/t/a* basal series observed in this study may not be maintained indefinitely. Longitudinal studies are, therefore, necessary to determine whether the Scott, Foresman program would result in an increased pace of achievement and would exceed that of children who learned in the more analytic procedures utilized in the Lippincott and *Early-to-Read i/t/a* series. Consequently, at the end of three years of instruction, the Scott, Foresman group's achievement may even surpass the achievement of the children in the other two basal series approaches. Before definitive statements indicating the superiority of the Lippincott and *Early-to-Read i/t/a* approaches can be made, a longitudinal study comparing these systems to Scott, Foresman must be accomplished. If the

Scott, Foresman group would equal, but not surpass, those children taught in the other basal series at the end of three years, studies of ancillary factors in such areas as personality characteristics would have to be made to make a decision as to whether greater success at the first-grade level has any other desirable side effects.

APPENDICES

APPENDIX A

TEACHER'S LOG

TEACHER'S NAME____________________

SCHOOL____________________

DISTRICT____________________

WEEK OF____________________

MONDAY				TUESDAY				WEDNESDAY				THURSDAY				FRIDAY			
TIME	O R G	ACTIVITY or SKILL	MAT'L	TIME	O R G	ACTIVITY or SKILL	MAT'L	TIME	O R G	ACTIVITY or SKILL	MAT'L	TIME	O R G	ACTIVITY or SKILL	MAT'L	TIME	O R G	ACTIVITY or SKILL	MAT'L

SAMPLE ENTRY: 9:00 - 9:30 - I *G a

9:00 - 9:30 II R b

NOTE: Place * before each activity or skill entry that is taught by the teacher.

APPENDIX A cont.

TEACHER'S LOG OF READING INSTRUCTION AND SUPPORTIVE ACTIVITIES

Supportive Activities

A. Building experiential background (pictures, discussion, trips, etc. as related to reading readiness program)

B. Oral reading not from basal

C. Choral speaking or reading

D. Concept development (language enrichment)

E. Experience stories (work-type or language charts)

F. Drama and puppetry

G. Story-telling, poetry

H. Spelling

I. Creative writing

J. Book reporting

K. Free play (classroom construction activities, puzzles, etc.)

L. Oral language development ("show and tell," retelling a story, speech instruction, etc.)

M. Art expression (related to a reading lesson or activity)

N. Recreational reading (trade books, supplementary readers, use of classroom or school library)

O. Handwriting

P. Reading in content areas (social studies, science, math.)

Reading Instruction

Teaching a reading lesson from the basal reader:
1. Preparation for reading a selection
2. Introducing new words
3. Silent reading
4. Oral reading

APPENDIX A cont.

5. Comprehension and interpretation
6. Combination of 1 through 5

Word analysis
7. Phonics
8. Structural analysis
9. Use of context clues
10. Combination of 7 and 8
11. Combination of 8 and 9
12. Sound blending
13. Dictionary (alphabetizing, etc.)

Vocabulary and comprehension
14. Vocabulary development
15. Resource and reference reading
16. Skills reinforcement (word recognition, comprehension)

Reading readiness skills
17. Listening
18. Following directions
19. Motor coordination
20. Visual discrimination
21. Auditory discrimination
22. Picture interpretation
    a. Basal reader series
    b. Workbooks (basal)
    c. Workbooks (not basal)
    d. Teacher-made materials ([illegible], vocabulary cards, picture illustrations, etc.)
    e. Commercial games and devices (flash cards, sound-symbol cards)
    f. Films, filmstrips
    g. Recordings (tape, phonograph)
    h. Newspapers and magazines (e.g., Weekly Reader)
    i. Trade books - T.O.
    j. Trade books - i/t/a
    k. Arts and crafts supplies
    l. Play equipment
    m. Flannel board
    n. Texts, pamphlets, etc. (social studies and science)
    o. Language arts text

I = Group
II = Group
III = Group

(Use Roman numerals IV, V, etc. for designating additional groups)

Ind = Individual
WC = Whole Class
SG = Children working in pairs or small groups

APPENDIX B

TEACHER RATING SCALE AND CRITERIA

Class Structure

0. Teacher is generally vague and directions seem confusing to the children.

1. There is generally more vagueness than clarity and more looseness than structure.

2. There is a moderate degree of structure and information on expectations. Some degree of vagueness.

3. Teacher is generally well organized and clear in assigning tasks -- directions and expectations clear, but not spelled out as below.

4. Teacher structures for the children -- gives detailed clear directions, and expectations are clearly spelled out in detail.

Extent of Class Participation

0. Class is generally unresponsive with only a very few children actually participating.

1. Participation is highly variable, but tends to be low quite often.

2. Teacher has a group of children who are participating well most of the time, but a fairly large group who are not consistently with the teacher.

3. Moderately high participation on part of most children, most of the time, but with some variability.

4. High participation on part of most children at all times.

Awareness of and Attention Paid to Individual Needs of Pupils

0. Total lack of awareness on the part of the teacher to the individual instructional needs of the pupils.

1. Limited awareness of and attention paid to individual needs of pupils.

2. There is moderate awareness and adjustment of instruction by the teacher based upon the needs of individual pupils in the class.

3. Teacher is generally aware of pupil needs and attempts to make the necessary instructional adjustments in light of these needs.

4. Teacher exceptionally aware of pupil needs with effective adjustment of instruction in light of these needs.

Overall Teacher Competence

0. Incompetent

1. Poor

2. Adequate

3. Good

4. Excellent

APPENDIX C

WORD RECOGNITION TESTS ADMINISTERED TO RANDOM SAMPLE:

GATES WORD PRONUNCIATION TEST

GARLAND PHONEMIC WORD TEST

PHONETICALLY REGULAR WORDS ORAL READING TEST

GATES WORD PRONUNCIATION TEST

EXAMINER'S COPY

DIRECTIONS: Have the child read the words out loud. Tell him you would like him to read som
words for you. If he fails the first time, ask him to try the word again. Continue un
til ten consecutive words have been missed. As the words become difficult, specia
care should be taken to encourage the child. The score is one point for each wor
correctly pronounced on the first trial, one-half point for each word correctly pro
nounced on the second trial. (Note: 9½ correct would be scored as 10.)

---

1. so
2. we
3. on
4. go
5. the
6. out
7. how
8. may
9. king
10. here
11. grow
12. late
13. every
14. about
15. paper
16. blind
17. window
18. family
19. perhaps
20. plaster
21. passenger
22. wonder
23. interest
24. chocolate
25. dispute
26. portion
27. conductor
28. brightness
29. intelligent
30. construct
31. position
32. profitable
33. irregular
34. schoolmaster
35. lamentation
36. community
37. satisfactory
38. illustrious
39. superstition
40. affectionate

Child's name: ______________________ Test date ___ ___ ___

Examiner: ______________________ Birth date ___ ___ ___

Age: ________

KARLSEN PHONEMIC WORD TEST

EXAMINER'S COPY

DIRECTIONS:
1. Hand the PUPIL'S COPY to the pupil.
2. Say: "Read these words out loud."
3. Note the pupil's errors on this sheet.
4. Do not give the pupil a second chance, but accept immediate self-correction.
5. Continue until the child misses 5 consecutive words.
6. The score is the number of words pronounced correctly.

---

1. fit
2. tap
3. red
4. get
5. would
6. mother
7. down
8. age
9. think
10. long
11. kind
12. yard
13. foot
14. gold
15. freeze
16. chair
17. mouth
18. carry
19. hope
20. beat
21. leaf
22. cowboy
23. furniture
24. page
25. push
26. huge
27. snowball
28. thirteen
29. scare
30. sunshine
31. gymnasium
32. join
33. usual
34. tomb
35. teaspoon
36. monument
37. senior
38. flute
39. behave
40. faucet

Child's name: ______________________ Test date ___ ___ ___

Examiner: ______________________ Birth date ___ ___ ___

Age: ______

PHONETICALLY REGULAR WORDS ORAL READING TEST

Child's Name____________________Date_________________________

School_____________________Room________Code Number__________

Examiner__________________Number of words read correctly_____

1. nap
2. pen
3. hid
4. job
5. rug
6. shade
7. drive
8. joke
9. mule
10. plain
11. hay
12. keen
13. least
14. loan
15. show
16. walk
17. haul
18. jaw
19. soil
20. joy
21. frown
22. trout
23. term
24. curl
25. birch
26. rare
27. star
28. porch
29. smooth
30. shook

Directions: Have pupil read words from one copy while examiner makes another copy. Do not give pupil a second chance but accept immediate self-correction. Let every student try the whole first column. If he gets two words correct from word number six on, let him try the whole second column.

APPENDICES D, E, F

CORRELATION MATRICES OF THE READING READINESS TESTS AND THE STANFORD WORD READING, PARAGRAPH MEANING, AND WORD STUDY SKILLS TESTS FOR THE LIPPINCOTT, i/t/a, AND SCOTT, FORESMAN GROUPS

Variable 1. Murphy-Durrell Phonemes

2. Murphy-Durrell Capital Letter Names
3. Murphy-Durrell Lower Case Letter Names
4. Murphy-Durrell Letter Names (Total)
5. Murphy-Durrell Learning Rate
6. Thurstone Pattern Copying
7. Thurstone Identical Forms
8. Metropolitan Word Meaning
9. Metropolitan Listening
10. Metropolitan Matching
11. Metropolitan Numbers
12. Metropolitan Copying
13. Metropolitan Alphabet
14. Metropolitan Total Score
15. Stanford Word Reading
16. Stanford Paragraph Meaning
17. Stanford Word Study Skills
18. Pitner-Cunningham IQ Score

APPENDIX D

CORRELATION MATRIX OF THE READING READINESS TESTS AND THE STANFORD WORD READING, PARAGRAPH MEANING AND WORD STUDY SKILLS TESTS FOR THE LIPPINCOTT TREATMENT GROUP

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.00	.59	.63	.63	.71	.52	.41	.42	.40	.43	.48	.41	.55	.60	.71	.63	.68	.53
2		1.00	.89	.97	.48	.28	.21	.26	.26	.29	.33	.24	.64	.43	.54	.52	.56	.40
3			1.00	.97	.61	.32	.25	.27	.27	.30	.36	.25	.67	.46	.58	.58	.58	.41
4				1.00	.56	.31	.24	.24	.27	.30	.37	.25	.67	.47	.58	.57	.59	.42
5					1.00	.40	.32	.26	.29	.33	.34	.25	.49	.46	.64	.62	.63	.44
6						1.00	.36	.29	.17	.33	.31	.50	.31	.42	.40	.37	.47	.47
7							1.00	.31	.26	.29	.33	.34	.28	.38	.32	.29	.29	.42
8								1.00	.52	.38	.53	.40	.43	.72	.31	.27	.28	.36
9									1.00	.43	.54	.47	.43	.73	.39	.42	.36	.43
10										1.00	.47	.54	.45	.72	.41	.40	.38	.40
11											1.00	.45	.55	.78	.42	.43	.40	.37
12												1.00	.46	.70	.39	.40	.32	.43
13													1.00	.74	.50	.54	.48	.37
14														1.00	.54	.52	.49	.53
15															1.00	.79	.83	.52
16																1.00	.75	.48
17																	1.00	.50
18																		1.00

APPENDIX E

CORRELATION MATRIX OF THE READING READINESS TESTS AND THE STANFORD WORD READING, PARAGRAPH MEANING AND WORD STUDY SKILLS TESTS FOR THE i/t/a TREATMENT GROUP

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.00	.55	.58	.59	.38	.49	.22	.26	.32	.37	.51	.39	.52	.61	.63	.61	.57	.42
2		1.00	.83	.96	.43	.41	.26	.20	.16	.26	.59	.36	.82	.66	.50	.54	.50	.35
3			1.00	.95	.48	.45	.19	.23	.25	.32	.62	.31	.84	.70	.58	.61	.54	.34
4				1.00	.47	.45	.23	.22	.21	.30	.63	.35	.35	.71	.56	.60	.54	.35
5					1.00	.39	.27	.28	.33	.32	.46	.30	.45	.54	.38	.39	.36	.34
6						1.00	.32	.23	.36	.40	.46	.62	.39	.61	.50	.45	.51	.45
7							1.00	.24	.27	.26	.34	.36	.18	.41	.31	.27	.32	.43
8								1.00	.46	.13	.37	.25	.25	.57	.26	.24	.25	.33
9									1.00	.30	.35	.29	.25	.60	.36	.30	.33	.29
10										1.00	.45	.38	.29	.58	.43	.45	.44	.43
11											1.00	.40	.59	.82	.54	.58	.55	.45
12												1.00	.31	.65	.42	.36	.40	.44
13													1.00	.73	.49	.49	.46	.30
14														1.00	.64	.62	.62	.55
15															1.00	.85	.79	.47
16																1.00	.76	.49
17																	1.00	.45
18																		1.00

APPENDIX F

CORRELATION MATRIX OF THE READING READINESS TESTS AND THE STANFORD WORD READING, PARAGRAPH MEANING AND WORD STUDY SKILLS TESTS FOR THE SCOTT, FORESMAN TREATMENT GROUP

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.00	.38	.44	.43	.46	.44	.25	.43	.36	.32	.47	.41	.39	.57	.51	.38	.56	.54
2		1.00	.86	.97	.37	.25	.17	.38	.21	.36	.42	.28	.82	.62	.55	.53	.54	.38
3			1.00	.96	.40	.25	.19	.40	.25	.37	.40	.32	.58	.65	.63	.58	.59	.42
4				1.00	.40	.26	.19	.41	.24	.38	.43	.31	.87	.66	.61	.57	.59	.41
5					1.00	.32	.23	.34	.20	.27	.29	.30	.41	.46	.38	.42	.39	.27
6						1.00	.23	.33	.34	.46	.45	.66	.27	.59	.38	.32	.43	.50
7							1.00	.22	.31	.32	.42	.26	.25	.44	.18	.27	.25	.37
8								1.00	.46	.39	.46	.33	.37	.68	.46	.39	.36	.42
9									1.00	.31	.36	.26	.26	.58	.26	.23	.32	.44
10										1.00	.49	.54	.41	.74	.41	.40	.40	.41
11											1.00	.50	.41	.78	.51	.46	.46	.48
12												1.00	.31	.69	.36	.31	.37	.54
13													1.00	.70	.57	.54	.55	.41
14														1.00	.63	.57	.61	.64
15															1.00	.78	.76	.43
16																1.00	.70	.36
17																	1.00	.46
18																		1.00

APPENDIX G

PARTIAL CORRELATIONS BETWEEN PREDICTORS: MURPHY-DURRELL PHONEMES, MURPHY-DURRELL LETTER NAMES, METROPOLITAN TOTAL SCORE, AND READING ACHIEVEMENT CRITERIA; STANFORD WORD READING (WR), PARAGRAPH MEANING (PM), AND WORD STUDY SKILLS (WSS) FOR LIPPINCOTT BOYS, GIRLS AND TOTAL GROUP

		Phonemes			Letter Names			Metropolitan Total		
		r	F	Sig.	r	F	Sig.	r	F	Sig.
Total Group	WR	.60	105.7	.01	.46	50.5	.01	.37	29.6	.01
	PM	.51	63.9	.01	.46	50.5	.01	.36	28.1	.01
	WSS	.57	90.1	.01	.48	56.3	.01	.30	19.1	.01
Boys	WR	.50	35.6	.01	.44	24.7	.01	.34	13.2	.01
	PM	.45	26.1	.01	.52	38.6	.01	.45	26.1	.01
	WSS	.52	39.7	.01	.56	47.1	.01	.27	8.30	.05
Girls	WR	.72	84.5	.01	.49	24.9	.01	.41	16.3	.01
	PM	.59	42.4	.01	.36	11.7	.01	.26	5.69	.01
	WSS	.63	50.9	.01	.38	13.5	.01	.35	10.8	.01

APPENDIX H

PARTIAL CORRELATIONS BETWEEN PREDICTORS: MURPHY-DURRELL PHONEMES, MURPHY-DURRELL LETTER NAMES, METROPOLITAN TOTAL SCORE, AND READING ACHIEVEMENT CRITERIA; STANFORD WORD READING (WR), PARAGRAPH MEANING (PM), AND WORD STUDY SKILLS (WSS) FOR i/t/a BOYS, GIRLS AND TOTAL GROUP

		Phonemes			Letter Names			Metropolitan Total		
		r	F	Sig.	r	F	Sig.	r	F	Sig.
Total Group	WR	.54	95.5	.01	.48	71.2	.01	.51	82.9	.01
	PM	.51	83.9	.01	.52	88.1	.01	.48	70.7	.01
	WSS	.47	67.5	.01	.46	61.7	.01	.50	77.2	.01
Boys	WR	.51	42.5	.01	.47	33.1	.01	.55	52.8	.01
	PM	.57	57.4	.01	.52	45.1	.01	.54	49.2	.01
	WSS	.46	32.8	.01	.41	23.9	.01	.50	39.1	.01
Girls	WR	.58	56.1	.01	.51	40.0	.01	.44	26.7	.01
	PM	.43	25.2	.01	.51	40.0	.01	.39	20.2	.01
	WSS	.48	32.7	.01	.51	39.6	.01	.49	35.5	.01

APPENDIX I

PARTIAL CORRELATIONS BETWEEN PREDICTORS: MURPHY-DURRELL PHONEMES, MURPHY-DURRELL LETTER NAMES, METROPOLITAN TOTAL SCORE, AND READING ACHIEVEMENT CRITERIA: STANFORD WORD READING (WR), PARAGRAPH MEANING (PM), AND WORD STUDY SKILLS (WSS) FOR SCOTT, FORESMAN BOYS, GIRLS AND TOTAL GROUP

		Phonemes			Letter Names			Metropolitan Total		
		r	F	Sig.	r	F	Sig.	r	F	Sig.
Total Group	WR	.37	36.2	.01	.53	88.0	.01	.51	79.4	.01
	PM	.24	13.8	.01	.50	75.1	.01	.48	65.9	.01
	WSS	.42	48.9	.01	.49	71.6	.01	.46	59.4	.01
Boys	WR	.30	11.35	.01	.50	39.5	.01	.50	37.4	.01
	PM	.16	3.15	-	.51	41.4	.01	.49	37.2	.01
	WSS	.30	11.1	.01	.48	34.7	.01	.41	23.8	.01
Girls	WR	.44	26.3	.01	.55	46.5	.01	.53	42.8	.01
	PM	.31	11.23	.05	.47	29.9	.01	.46	29.4	.01
	WSS	.56	50.38	.01	.49	34.5	.01	.50	36.2	.01

REFERENCES

Bremer, N. Do readiness tests predict success in reading? *Elementary School Journal*, 1959, 59, 222-224

Carrol, M. W. Sex differences in reading readiness at the first-grade level. *Elementary English*, 1948, 25, 370-375.

Dean, C. D. Predicting first-grade reading achievement. *Elementary School Journal*, 1939, 609-616.

Deputy, E. C. Predicting first-grade reading achievement. Columbia University Contributions to Education, No. 426. New York: Bureau of Publications, Teachers College, Columbia University, 1930.

Dewey, G. *Relativ frequency of English speech sounds*. Cambridge: Harvard University Press, 1923.

Downing, J. A. *The i.t.a. reading experiment*. London: Evans Brothers, 1964. (a)

Downing, J. A. The i.t.a. reading experiment. *The Reading Teacher*, 1964, 18, 105-110. (b)

Gage, N. L. (Ed.) *Handbook of research on teaching*. Chicago: Rand McNally, 1963.

Guilford, J. P. *Fundamental statistics in psychology and education*. (4th ed.) New York: McGraw-Hill, 1965.

Harris, A. J. *How to increase reading ability*. New York: Longmans Green, 1961.

Heilman, A. W. *Principles and practices of teaching reading*. Columbus: Charles E. Merrill Books, 1961.

Hotelling, H. The most predictable criterion. *Journal of Educational Psychology*, 1935, 26, 139-142.

Karlin, R. The prediction of reading success and reading readiness tests. *Elementary English*, 1957, 34, 320-322.

Lee, J. M., Clark, W. W., & Lee, D. M. Measuring reading readiness. *Elementary School Journal*, 1934, 34, 656-666.

Mazurkiewicz, A. J. Lehigh - Bethlehem - i/t/a study - interim report six. *The Journal of the Reading Specialist*, 1964, 4, 3-6. (a)

Mazurkiewicz, A. J. Teaching reading in America using the initial teaching alphabet. *Elementary English*, 1964, 41, 766-772. (b)

Nila, Sister Mary. Foundations of a successful reading program. Education, 1953, 73, 543-555.

Pitman, I. J. Learning to read: an experiment. Journal of the Royal Society of Arts. London, 1961, 109, 149-180, 196.

Prescott, G. A. Sex differences in metropolitan readiness test results. Journal of Educational Research, 1955, 48, 605-610.

Samuels, F. Sex differences in reading achievement. Journal of Educational Research, 1943, 36, 594-603.

Scheffe, H. The analysis of variance. New York: John Wiley and Sons, 1959.

Spache, G. D. A new readability formula for primary grade materials. Elementary School Journal, 1953, 53, 410-413.

Torgerson, T. L. & Adams, G. S. Measurement and evaluation for the elementary-school teacher. New York: The Dryden Press, 1954.

Wilson, F. T., Burke, A., & Fleming, C. W. Sex differences in beginning reading in a progressive school. Journal of Educational Research, 1939, 32, 570-582.

Wilson, F. T. & Fleming, C. W. Correlations of chronological age, mental age and intelligence quotient with other abilities and traits in Grade I. Journal of Genetic Psychology, 1937, 50, 323-337.